# On the Capacity of Channels with Synchronization Errors

Aravind R. Iyengar, Paul H. Siegel, *Fellow, IEEE* and Jack K. Wolf, *Life Fellow, IEEE*

***Abstract*—We consider a new formulation of a class of synchronization error channels and derive analytical bounds and numerical estimates for the capacity of these channels. For the binary channel with only deletions, we obtain an expression for the symmetric information rate in terms of subsequence weights which reduces to a tight lower bound for small deletion probabilities. We are also able to exactly characterize the Markov-1 rate for the binary channel with only replications. For a channel that introduces deletions as well as replications of input symbols, we design two sequences of approximating channels with favourable properties. In particular, we parameterize the state space associated with these approximating channels and show that the information rates approach that of the deletion-replication channel as the state space grows. For the case of the channel where deletions and replications occur with the same probabilities, a stronger result in the convergence of mutual information rates is shown. The numerous advantages this new formulation presents are explored.**

***Index Terms*—Synchronization errors, deletions, insertions, replications, channel capacity.**

## I. Introduction

CHANNELS with synchronization errors have been familiar to information and coding theorists and practitioners alike ever since the advent of the digital information era. Although Dobrushin [2] established the coding theorem for such channels as early as 1967, tackling these channels in terms of estimating information rates and constructing codes with good performance have proved to be very tough. In the last decade, significant progress has been made in estimating achievable information rates for certain channels with synchronization errors. However, a coding scheme with provably "good" performance remains elusive thus far.

In this paper, we start with Dobrushin's model of channels with synchronization errors, henceforth referred to as the *synchronization error channel* (SEC), and convert it into an equivalent channel with states. Using this alternative model, we construct a sequence of channels that "approximate" the SEC and whose limit is the SEC. We use these approximate channels to derive some results about information rates achievable over the SEC. Although the motivation behind the alternative model is straightforward, its use to obtain non-trivial bounds on the capacity of the SEC has, to the best of our knowledge, not been found in literature. While the present paper concerns only a few asymptotic results on information rates of the SEC, we think that the model presented here can be utilized to design codes for SECs in general.

The remainder of this paper is organized as follows. In Section II, we revisit Dobrushin's model of an SEC and recall the main results on capacity of SECs. Through much of the paper, we consider a special case of the generic SEC—the deletion, replication channel (DRC)—and construct an equivalent channel by viewing the DRC as a channel with states in Section III. Under further special cases of channels with only deletions or only replications, we give some simple, non-trivial and sometimes tight bounds on the capcity in Sections IV-A and IV-B. We then construct a sequence of finite state channels that approximate the DRC and establish certain properties of this sequence of channels that serve as bounds for the capacity of the DRC in Section V-A. In Section VI, we note the application of similar strategies to more general SECs, and we conclude with summary and remarks in Section VII.

A. R. Iyengar was with the Department of Electrical and Computer Engineering and the Center for Magnetic Recording Research, University of California, San Diego. He is now with Qualcomm Technologies Inc., Santa Clara CA 95051 USA (e-mail: ariyengar@qti.qualcomm.com). P. H. Siegel is with the Department of Electrical and Computer Engineering and the Center for Magnetic Recording Research, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: psiegel@ucsd.edu). J. K. Wolf (deceased) was with the Department of Electrical and Computer Engineering and the Center for Magnetic Recording Research, University of California, San Diego, La Jolla, CA 92093 USA.

This work was supported in part by the Center for Magnetic Recording Research and by the National Science Foundation under the Grant CCF-0829865. A summary of some of the results in Sections II through IV was presented at the 2011 International Symposium on Information Theory (ISIT), St. Petersburg, Russia [1].

## II. Synchronization Error Channels

*Remark 1 (Notation):* Non-random variables are written as lowercase letters, e.g., $n$. We denote sets by double-stroke uppercase letters, e.g., $\mathbb{X}$. We will reserve $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$ to denote the sets of natural numbers, integers and real numbers, respectively. $\mathbb{Z}^+$ denotes the set of non-negative integers. We define

$$\begin{aligned}
[n] &\triangleq \{1,2,\cdots,n\}, n\in\mathbb{N},\\
[0] &\triangleq \emptyset,\\
[m:n] &\triangleq \begin{cases}\{m,m+1,\cdots,n\}, & m\le n,\\ \emptyset, & n<m.\end{cases} \text{ and}\\
\mathbb{Z}_{\pm m} &\triangleq \{-m,-m+1,\cdots,0,1,\cdots,m\}\ \forall\ m\in\mathbb{Z}^+.
\end{aligned}$$

For some $n\in\mathbb{N}$, we will let $\mathbb{X}^n$ denote the set of vectors of dimension $n$ with elements from $\mathbb{X}$. We will write $\overline{x}$ to denote a string, and $\lambda$ to denote the *empty string*. The *length* of a string, denoted $|x|$, is the number of symbols in it, and by definition, $|\lambda|=0$. With some abuse of notation, we will use "vectors of dimension $n$" and "strings of length $n$" interchageably. The set of all strings of length $n$ over the alphabet $\mathbb{X}$ is hence also

denoted $\mathbb{X}^n$, and $\mathbb{X}^0 = \{\lambda\}$. We write $\overline{\mathbb{X}}$ to denote the set of all strings over the set $\mathbb{X}$, i.e.,

$$\overline{\mathbb{X}} = \bigcup_{i=0}^{\infty} \mathbb{X}^i.$$

The bar "$\overline{\cdot}$" will denote the concatenation operation, so that $\overline{\overline{x} \cdot \overline{y}}$ is the concatenation of strings $\overline{x}$ and $\overline{y}$.

Throughout the paper, we assume an underlying probability space $(\mathbb{S}, \mathscr{B}, \mathsf{P})$ over which random variables, denoted by uppercase letters, e.g., $X$, are defined. Random vectors are denoted by uppercase letters with the *multiset* of indices as subscripts, e.g., $X_{[n]} = (X_1, X_2, \cdots, X_n)$, or $X_{Y_{[n]}}$ when the multiset of indices is itself the elements of a random vector $Y_{[n]}$. Random processes (assumed discrete-time) are denoted by script letters $\mathcal{X}$, or subscripted by the set of natural numbers, $X_{\mathbb{N}}$.

We will use the asymptotic notations $O(\cdot)$, $o(\cdot)$, $\omega(\cdot)$ as in [3], [4]. □

We start by defining the synchronization error channels as considered by Dobrushin [2].

*Definition 1 (Memoryless SECs):* Let $\mathbb{X}$ be a finite set. A *memoryless* synchronization error channel is specified by a stochastic matrix

$$\{q(\overline{y}|x), \overline{y} \in \overline{\mathbb{Y}}, x \in \mathbb{X}\}$$

where $\mathbb{Y}$ is the output alphabet. From the properties of a stochastic matrix, we have

$$0 \le q(\overline{y}|x) \le 1, \qquad \sum_{\overline{y} \in \overline{\mathbb{Y}}} q(\overline{y}|x) = 1 \ \forall \ x \in \mathbb{X}. \tag{1}$$

Further, we will assume that the mean value of the length of the output string arising from one input symbol is strictly positive and finite, i.e.,

$$0 < \sum_{\overline{y} \in \overline{\mathbb{Y}}} |\overline{y}| q(\overline{y}|x) < \infty. \tag{2}$$

For $x_{[n]} = (x_1, x_2, \cdots, x_n) \in \mathbb{X}^n$ and $\overline{y}_{[n]} = (\overline{y}_1, \overline{y}_2, \cdots, \overline{y}_n) \in \overline{\mathbb{Y}}^n$, we write

$$q_n(\overline{y}_{[n]}|x_{[n]}) = \prod_{i=1}^{n} q(\overline{y}_i|x_i).$$

Let $\overline{\overline{y}_{[n]}}$ denote the concatenation of strings $\overline{y}_i, i \in [n]$. Then the transition probabilities of the memoryless SEC are defined as

$$Q_n(\overline{y}|x_{[n]}) = \sum_{\overline{\overline{y}_{[n]}} = \overline{y}} q_n(\overline{y}_{[n]}|x_{[n]}) \tag{3}$$

for $\overline{y} \in \overline{\mathbb{Y}}$ and $x_{[n]} \in \mathbb{X}^n$. The memoryless SEC is given by the triplet $\mathbf{Q}_n \triangleq (\mathbb{X}, Q_n, \mathbb{Y})$, the input and the output alphabets, and the transition probabilities between input strings of length $n$ and all output strings. □

Consider the sequence of memoryless SECs $\{\mathbf{Q}_n\}_{n=1}^{\infty}$. Then, we have the following.

*Theorem 2 (Capacity [2]):* Let $X_{[n]}$ and $\overline{Y}$ denote the input and the output of the SEC $\mathbf{Q}_n$. Let

$$C_n = \sup_{\mathsf{P}(X_{[n]})} \frac{1}{n} I(X_{[n]}; \overline{Y}).$$

Then,

$$C = \lim_{n \to \infty} C_n = \inf_{n \ge 1} C_n$$

exists and is equal to the capacity of the sequence of SECs. ■

The quantity $C$ represents the maximum rate at which information can be transferred over the SEC with vanishing error probability. Furthermore, the following result shows that, in estimating the capacity of the SEC, we can restrict ourselves to a subclass of possible input processes $\mathcal{X}$.

*Proposition 3 (Markov Capacity [2]):* Let $\mathcal{X}_{\mathcal{M}}$ be a stationary, ergodic, Markov process over $\mathbb{X}$. Then the capacity of the sequence $\{\mathbf{Q}_n\}_{n=1}^{\infty}$ is

$$C = \sup_{\mathcal{X}_{\mathcal{M}}} \lim_{n \to \infty} \frac{1}{n} I(X_{[n]}; \overline{Y}).$$

The capacity is therefore the supremum of the rates achievable through stationary, ergodic, Markov processes $\mathcal{X}_{\mathcal{M}}$. ■

We will now give an example of a memoryless SEC. Throughout the paper, we will assume that the input alphabet for the SECs is $\mathbb{X} = \{0, 1\}$, i.e., the channels considered are *binary* memoryless SECs. However, we note here that all the results in the paper can be straightforwardly extended to the case where $\mathbb{X}$ is any finite set.

*Example 4 (Deletion-Replication Channel (DRC)):* Consider the binary SEC with $\mathbb{X} = \mathbb{Y} = \{0, 1\}$ and the following stochastic matrix.

$$q(\overline{y}|x) = \begin{cases} p_{\mathsf{d}}, & \overline{y} = \lambda \\ p_{\mathsf{t}} p_{\mathsf{r}}^{\ell-1}, & \overline{y} = x^{\ell}, \ \forall \ \ell \ge 1. \end{cases}$$

Intuitively, we can think of $p_{\mathsf{d}}$ as the deletion probability, $p_{\mathsf{t}}$ as the transmission probability, and $p_{\mathsf{r}}$ as the replication probability, i.e., when $x \in \mathbb{X}$ is sent, it is either deleted with probability $p_{\mathsf{d}}$, or transmitted and replicated $(\ell - 1)$ times with probability $p_{\mathsf{t}} p_{\mathsf{r}}^{\ell-1}$ for $\ell \ge 1$. From (1), we get for $p_{\mathsf{r}} < 1$

$$p_{\mathsf{d}} + \sum_{\ell=1}^{\infty} p_{\mathsf{t}} p_{\mathsf{r}}^{\ell-1} = p_{\mathsf{d}} + \frac{p_{\mathsf{t}}}{1 - p_{\mathsf{r}}} = 1,$$

or equivalently

$$p_{\mathsf{t}} = (1 - p_{\mathsf{d}})(1 - p_{\mathsf{r}}). \tag{4}$$

From (2),

$$0 < \sum_{\ell=1}^{\infty} \ell p_{\mathsf{t}} p_{\mathsf{r}}^{\ell-1} = \frac{p_{\mathsf{t}}}{(1 - p_{\mathsf{r}})^2} = \frac{1 - p_{\mathsf{d}}}{1 - p_{\mathsf{r}}} < \infty$$

where we use Equation (4). Hence $(p_{\mathsf{d}}, p_{\mathsf{r}}) \in [0, 1)^2$. Note that when $p_{\mathsf{r}} = 0$, the DRC is the same as the *binary deletion channel* (BDC); and when $p_{\mathsf{d}} = 0$, it is the *binary replication channel* (BRC), also referred to as the *geometric binary sticky channel* [5]. □

$$\begin{aligned}P_n(\overline{y}|x_{[n]}, Z_0=0) &= \sum_{\{\overline{z}:|\overline{z}|=|\overline{y}|\}} \mathsf{P}(\overline{Z}=\overline{z}, \overline{Y}=\overline{y}|X_{[n]}=x_{[n]}, Z_0=0)\\ &= \sum_{\{\overline{z}:|\overline{z}|=|\overline{y}|\}} \mathsf{P}(\overline{Z}=\overline{z}|Z_0=0)\mathsf{P}(\overline{Y}=\overline{y}|X_{[n]}=x_{[n]}, Z_0=0, \overline{Z}=\overline{z})\\ &= \sum_{\{\overline{z}:|\overline{z}|=|\overline{y}|\}} \prod_{i=1}^{|\overline{y}|}\Big(\mathsf{P}(Z_i=z_i|Z_{i-1}=z_{i-1}, Z_0=0)\mathsf{P}(Y_i=y_i|X_{[n]}=x_{[n]}, Z_i=z_i)\Big)\\ &= \sum_{\{\overline{z}:|\overline{z}|=|\overline{y}|\}} \prod_{i=1}^{|\overline{y}|}\Big(\mathsf{P}(Z_i=z_i|Z_{i-1}=z_{i-1})\mathbb{1}_{\{y_i=x_{i-z_i}\}}\Big).\end{aligned} \tag{5}$$

The BDC has been the most well-studied SEC. In [6], the author surveys the results that were known prior to 2009. To summarize, the best known lower bounds were obtained, chronologically, through bounds on the cutoff rate for sequential decoding [7], bounding the rate with a first-order Markov input [8], reduction to a Poisson-repeat channel [9], analyzing a "jigsaw-puzzle" coding scheme [10], or by directly bounding the information rate by analyzing the channel as a joint renewal process [11]. Recently, [12] and [13] independently gave the capacity of a BDC with small deletion probabilities, and showed that it is achieved by independent and uniformly distributed (i.u.d.) inputs. The known upper bounds for the BDC have been obtained by genie-aided decoder arguments [14], [15]. An idea from [15] was extended to obtain some analytical lower bounds on the capacity of channels that involve substitution errors as well as insertions or deletions [16]. The idea in [12] was extended to obtain a better approximtion for the capacity of the BDC with small deletion probabilities in [17].

In contrast to these existing results, our approach explicitly characterizes the achievable information rates in terms of "subsequence-weights", which is a measure relevant in ML decoding for the BDC [6]. Additionally, the method proposed here gives the tight bound on capacity for small deletion probabilities obtained in [12] more directly[1].

For the BRC, [5] obtained lower bounds on the capacity by numerically estimating the capacity per unit cost of the equivalent channel of runs through optimization of 8 and 16 bit codes. Here, we obtain direct analytical lower bounds on the capacity. These, to the best of our knowledge, represent the only analytical bounds for the capacity of the BRC. Moreover, we obtain an exact expression for the Markov-1 rate for the BRC which conclusively disproves the conjecture that the capacity of SECs is a convex function of the channel parameter.

We will use the DRC as a running example of an SEC. In Section VI, we discuss the extension of the results presented to a more general class of SECs.

## III. DRC AS A CHANNEL WITH STATES

We now construct a channel with states that is equivalent to the DRC introduced in Example 4. Dobrushin's model of SEC (cf. Definition 1) tracks the output string generated by each input symbol. In our model, we track the input symbol that gave rise to each output symbol.

### A. Channel Model

*Definition 5 (DRC with states):* For a fixed $n \in \mathbb{N}$, we write

$$Y_i = X_{\Gamma_i} = X_{i-Z_i}, i \in [N_n] \tag{6}$$

where $Z_i \in \mathbb{Z}$ is the "state" of the channel and

$$N_n \triangleq \sup\{i \geq 0 : \Gamma_i \leq n | \Gamma_0 = 0\}.$$

We will refer to the random variable $N_n$ as the *output length* corresponding to $n$ input symbols. The state process $\mathcal{Z}$ is independent of the channel input process, and is a first-order Markov process over the set of integers $\mathbb{Z}$ with transition probabilities for each $i \in \mathbb{N}$ given by

$$\mathsf{P}(Z_i = z_i|Z_{i-1}=z_{i-1}) = \begin{cases} p_{\mathsf{r}}, & z_i = z_{i-1}+1 \\ p_{\mathsf{t}}p_{\mathsf{d}}^{\ell}, & z_i = z_{i-1}-\ell,\ \forall\ \ell \geq 0, \end{cases} \tag{7}$$

where we define $p_{\mathsf{t}}$ as in Equation (4) assuming $(p_{\mathsf{d}}, p_{\mathsf{r}}) \in [0,1)^2$. We will refer to the process $\mathit{\Gamma} \triangleq \Gamma_{\mathbb{N}}$ where $\Gamma_i = i - Z_i$ as the *index process*.

We also assume the boundary condition that $Z_0 = \Gamma_0 = 0$, i.e., the channel is perfectly synchronized before transmission commences. Note that the transition probabilities in (7) indeed are well-defined since $\forall\ z_{i-1} \in \mathbb{Z}$, as $p_{\mathsf{d}} < 1$,

$$\sum_{z_i} p(z_i|z_{i-1}) = p_{\mathsf{r}} + \sum_{\ell=0}^{\infty} p_{\mathsf{t}}p_{\mathsf{d}}^{\ell} = p_{\mathsf{r}} + \frac{p_{\mathsf{t}}}{1-p_{\mathsf{d}}} = 1.$$

With the above definition, for $\overline{y} \in \overline{\mathbb{Y}}$ and $x_{[n]} \in \mathbb{X}^n$, the channel transition probabilities are given as in Equation (5). Note that in the terms within the parenthesis on the right hand side of Equation (5), the first term is completely specified by the transition probabilities (7) of the channel state process $\mathcal{Z}$, and the second term is 0 or 1 accordingly as $y_i \neq x_{i-z_i}$ or $y_i = x_{i-z_i}$ respectively.

For each $n \in \mathbb{N}$, we define the DRC with states as the channel $\mathbf{P}_n \triangleq (\mathbb{X}, P_n, \mathbb{Y})$. □

We will start by proving a few properties of the output length $N_n$ and the channel state $\mathcal{Z}$ and index processes $\mathit{\Gamma}$ which will

[1]Note that although we obtain the same lower bound for the capacity of the BDC as in [12], we do not prove a converse here.

be made use of subsequently. We will relegate the proofs of the following two lemmas to Appedices I and II respectively.

*Lemma 6 (Properties of $N_n$):* The output length $N_n$ satisfies the following properties:

(i) For any $n \in \mathbb{N}$, $N_n < \infty$ a.s..
(ii) $N_n \to \infty$ as $n \to \infty$ a.s..
(iii) $\frac{N_n}{n} \to \frac{1-p_\mathsf{d}}{1-p_\mathsf{r}}$ a.s. as $n \to \infty$. ■

*Lemma 7 (Properties of $\mathcal{Z}, \Gamma$):* The channel state process $\mathcal{Z}$ and the index process $\Gamma$ satisfy the following properties:

(i) $\mathcal{Z}$ and $\Gamma$ are first-order, time-homogeneous, shift-invariant Markov chains. Further, $\mathcal{Z}$ is irreducible and aperiodic.
(ii) $\Gamma$ is almost surely non-decreasing, i.e.,

$$\Gamma_{i+j} \geq \Gamma_i \ \forall \ j \geq 0, i \in \mathbb{N} \text{ a.s..}$$

For any $n \in \mathbb{N}$, a realization of $Z_{[n]}$ such that the corresponding $\Gamma_{[n]}$ realization satisfies the above monotonicity property is called a *compatible* state path.
(iii) For every $i \in \mathbb{N}$,

$$H(Z_i|Z_{i-1}) = H(\Gamma_i|\Gamma_{i-1}) = h_2(p_\mathsf{r}) + \frac{1-p_\mathsf{r}}{1-p_\mathsf{d}} h_2(p_\mathsf{d}),$$

where $h_2(x) \triangleq -x\log_2 x - (1-x)\log_2(1-x)$, for $x \in [0,1]$, is the *binary entropy function* [18]. Here, we assume from continuity that $0\log_2 0 = 0$. Consequently, for every $n \in \mathbb{N}$,

$$H(Z_{[n]}) = H(\Gamma_{[n]}) = n\Big(h_2(p_\mathsf{r}) + \frac{1-p_\mathsf{r}}{1-p_\mathsf{d}} h_2(p_\mathsf{d})\Big). \ ■$$

Note that the $\mathcal{Z}$ process is *not* stationary because we fix $Z_0 = 0$. The $\Gamma$ process is clearly not stationary since $\Gamma_i$ depends on $i$. From Lemma 7 (ii), we can show that for $1 \leq n \leq m < \infty$,

$$N_n \leq N_m \text{ a.s..} \tag{8}$$

*Proposition 8 (Channel Equivalence):* For each $n \in \mathbb{N}$, the channels $\mathbf{Q}_n$ and $\mathbf{P}_n$ are equivalent.

*Proof:* Both $\mathbf{Q}_n$ and $\mathbf{P}_n$ have the same input and output alphabets $\mathbb{X}$ and $\mathbb{Y}$, respectively. The correspondence between the transition probabilities $Q_n$ and $P_n$ in Equations (3) and (5) is evident by the following observations:

(i) For every parsing of $\overline{y} \in \overline{\mathbb{Y}}$ as $\overline{y}_{[n]}$ in Equation (3), there is a corresponding state path $\overline{z} \in \overline{Z}$ in Equation (5).
(ii) For every compatible state path $\overline{z} \in \overline{Z}$ in Equation (5) (See Lemma 7), there is a corresponding parsing of $\overline{y} \in \overline{\mathbb{Y}}$ in Equation (3).
(iii) For these corresponding parsings of $\overline{y}$ and compatible state paths $\overline{z}$, the terms within the parenthesis on the right hand side of Equation 5, when grouped according to the output symbols arising from the same input symbol, spell out exactly the same probability as the terms $q(\overline{y}_i|x_i)$.

Therefore, except on a set of zero probability (state paths that are not compatible), the probability measures $Q_n$ and $P_n$ are equal. This implies the equivalence of the channels $\mathbf{Q}_n$ and $\mathbf{P}_n$. ■

As a consequence of the above equivalence, the results of Theorem 2 and Proposition 3 carry forward to the sequence of channels $\{\mathbf{P}_n\}_{n=1}^{\infty}$ specified by Equations (6) and (7).

*Corollary 9 (Dobrushin's results for $\{\mathbf{P}_n\}_{n=1}^{\infty}$):* For input $X_{[n]}$ and output $Y_{[N_n]}$ of the channel $\mathbf{P}_n$, the quantity

$$\begin{aligned} C &= \lim_{n\to\infty} \sup_{\mathsf{P}(X_{[n]})} \frac{1}{n} I(X_{[n]}; Y_{[N_n]}) \\ &= \sup_{\mathcal{X}_\mathcal{M}} \lim_{n\to\infty} \frac{1}{n} I(X_{[n]}; Y_{[N_n]}), \end{aligned}$$

where $\mathcal{X}_\mathcal{M}$ represents stationary, ergodic, Markov processes over $\mathbb{X}$, exists and is equal to the capacity of the sequence of channels $\{\mathbf{P}_n\}_{n=1}^{\infty}$. ■

We will henceforth restrict our attention to this class of input processes. The following is a useful result whose proof is deferred to Appendix III.

*Proposition 10 (Stationarity):* The channel output process $\mathcal{Y}$ is stationary for stationary input processes $\mathcal{X}$. ■

As a consequence of the above result, the *entropy rate* $\mathcal{H}(\mathcal{Y})$ of the output process is well-defined [18].

### *B. Bounds on the Capacity of the DRC*

The formulation of the DRC as a channel with states allows us to immediately establish the following.

*Proposition 11 (Simple bounds on $C$):* For the DRC,

$$(1-p_\mathsf{d})\Big(1 - \frac{h_2(p_\mathsf{r})}{1-p_\mathsf{r}}\Big) - h_2(p_\mathsf{d}) \leq C \leq 1 - p_\mathsf{d}.$$

*Proof:* We can write

$$\begin{aligned} I(X_{[n]};Y_{[N_n]}) &= I(X_{[n]}; Y_{[N_n]}, Z_{[N_n]}) - I(X_{[n]}; Z_{[N_n]}|Y_{[N_n]}) \\ &\overset{(a)}{=} I(X_{[n]}; Y_{[N_n]}|Z_{[N_n]}) - I(X_{[n]}; Z_{[N_n]}|Y_{[N_n]}) \\ &\overset{(b)}{=} (1-p_\mathsf{d})H(X_{[n]}) - I(X_{[n]}; Z_{[N_n]}|Y_{[N_n]}), \end{aligned} \tag{9}$$

where $(a)$ is true because $\mathcal{X} \perp\!\!\!\perp \mathcal{Z}$ and $(b)$ from the fact that the DRC, given the $\mathcal{Z}$ process realization, is equivalent to a binary erasure channel (BEC) with erasure rate $p_\mathsf{d}$. Then,

$$n(1-p_\mathsf{d}) \geq I(X_{[n]}; Y_{[N_n]}) \geq (1-p_\mathsf{d})H(X_{[n]}) - H(Z_{[N_n]}).$$

From Lemma 7 (iii), and since, for any finite $n$, we have the extra knowledge that $Z_i \geq i - n$ by definition of $N_n$, we can show that

$$H(Z_{[N_n]}) \leq \mathsf{E}(N_n)\Big(h_2(p_\mathsf{r}) + \frac{1-p_\mathsf{r}}{1-p_\mathsf{d}} h_2(p_\mathsf{d})\Big).$$

Note that the extra information $Z_i \geq i - n$ becomes tautological when $n \to \infty$, and hence

$$\lim_{n\to\infty} \frac{H(Z_{[N_n]})}{n} = \Big(\lim_{n\to\infty} \frac{\mathsf{E}(N_n)}{n}\Big)\Big(h_2(p_\mathsf{r}) + \frac{1-p_\mathsf{r}}{1-p_\mathsf{d}} h_2(p_\mathsf{d})\Big).$$

From Lemma 6, and for independent uniformly distributed inputs, the claim follows. ■

Proposition 11 gives bounds on the capacity for $(p_\mathsf{d}, p_\mathsf{r}) \in [0,1)^2$. Three special cases of the DRC are of particular interest: the binary deletion channel (BDC) with $p_\mathsf{d} = p, p_\mathsf{r} =$

0; the *symmetric* deletion-replication channel (SDRC) with $p_\mathsf{d} = p_\mathsf{r} = p$; and the binary replication channel (BRC) with $p_\mathsf{d} = 0, p_\mathsf{r} = p$. Specializing Proposition 11 to these cases gives us the following results.

*Corollary 12 (Bounds on $C$ for special cases):* We have

$$\begin{aligned} 1 - p - h_2(p) &\leq C_{\text{BDC}} \leq 1 - p, \\ 1 - p - 2h_2(p) &\leq C_{\text{SDRC}} \leq 1 - p, \\ 1 - \frac{h_2(p)}{1-p} &\leq C_{\text{BRC}} \leq 1. \end{aligned} \qquad \blacksquare$$

Although the bounds in Corollary 12 have simple closed-form expressions with well known information theoretic functions, they are loose compared to the best known (analytical or numerical) bounds for the capacity of these channels. We can, however, improve these bounds. We have from Equation (9),

$$\begin{aligned} I(X_{[n]}; Y_{[N_n]}) = (1 - p_\mathsf{d})H(X_{[n]}) + I(Y_{[N_n]}; Z_{[N_n]}) \\ - H(Z_{[N_n]}) + H(Z_{[N_n]} | X_{[n]}, Y_{[N_n]}). \end{aligned} \tag{10}$$

Writing the entropy rate of the input process $\mathcal{X}$ as $\mathcal{H}(\mathcal{X})$ and defining

$$\hat{\mathcal{H}}(\mathcal{Z}) \triangleq \lim_{n\to\infty} \frac{H(Z_{[N_n]})}{n}, \quad \hat{\mathcal{H}}(\mathcal{Y}) \triangleq \lim_{n\to\infty} \frac{H(Y_{[N_n]})}{n},$$
$$\text{and} \quad \hat{\mathcal{H}}(\mathcal{Z}|\mathcal{X},\mathcal{Y}) \triangleq \lim_{n\to\infty} \frac{H(Z_{[N_n]}|X_{[n]}, Y_{[N_n]})}{n},$$

from Lemma 7 and Equation (10), we can bound

$$C \geq \sup_{\mathcal{X}} \Big( (1 - p_\mathsf{d})\mathcal{H}(\mathcal{X}) + \hat{\mathcal{H}}(\mathcal{Z}|\mathcal{X},\mathcal{Y}) \Big) - \frac{1 - p_\mathsf{d}}{1 - p_\mathsf{r}} h_2(p_\mathsf{r}) - h_2(p_\mathsf{d}).$$

*Lemma 13:* Let $H_n \triangleq \frac{1}{n} H(Z_{[N_n]}|X_{[n]}, Y_{[N_n]})$ for $n \in \mathbb{N}$. Then, for the sequence $\{H_n\}_{n=1}^{\infty}$,

$$\hat{\mathcal{H}}(\mathcal{Z}|\mathcal{X},\mathcal{Y}) = \lim_{n\to\infty} H_n = \sup_{n\geq 1} H_n. \qquad \blacksquare$$

The proof is given in Appendix IV. The above result implies that if we could evaluate (or lower bound) $H_n$ for some $n$, that could be used to estimate a lower bound on $C$.

*Proposition 14:* For the DRC,

$$C \geq \sup_{\mathcal{X}} \Big( \mathcal{H}(\mathcal{X}) + \frac{H(Z_1|\mathcal{X},\mathcal{Y})}{1 - p_\mathsf{r}} \Big)(1 - p_\mathsf{d}) - \frac{1 - p_\mathsf{d}}{1 - p_\mathsf{r}} h_2(p_\mathsf{r}) - h_2(p_\mathsf{d}).$$

*Proof:* We have

$$\begin{aligned} H_n &= \frac{1}{n} H(Z_{[N_n]}|X_{[n]}, Y_{[N_n]}) \\ &= \frac{1}{n} \mathsf{E}\Big( \sum_{i=1}^{N_n} H(Z_i | Z_{[i-1]} = z_{[i-1]}, X_{[n]}, Y_{[N_n]}) \Big) \\ &= \frac{1}{n} \mathsf{E}\Big( \sum_{i=1}^{N_n} H(Z_i | Z_{i-1} = z_{i-1}, X_{[i-1-z_{i-1}:n]}, Y_{[i:N_n]}) \Big) \end{aligned}$$

where the last equality follows from the conditional independence of $Z_i$ on $Z_{[i-2]}, X_{[i-2-Z_{i-1}]}$ and $Y_{[i-1]}$ given $Z_{i-1}$. From the time-homogeneity and shift-invariance of the $\mathcal{Z}$ process (See Lemma 7), as $n \to \infty$, the summand in the above expression

$$\begin{aligned} &H(Z_i | Z_{i-1} = z_{i-1}, X_{[i-1-z_{i-1}:n]}, Y_{[i:N_n]}) \\ &\quad \to H(Z_1 | Z_0 = 0, X_{\mathbb{N}}, Y_{\mathbb{N}}) \triangleq H(Z_1|\mathcal{X},\mathcal{Y}). \end{aligned}$$

Since $\frac{\mathsf{E}(N_n)}{n} \to \frac{1-p_\mathsf{d}}{1-p_\mathsf{r}}$, optimizing over input processes $\mathcal{X}$ gives us the desired result. $\blacksquare$

It is not easy to evaluate the bound in Proposition 14. However, we can further lower bound the capacity by introducing some conditioning.

*Lemma 15:* The sequence of lower bounds $\{D_i^{\mathcal{X}}\}_{i=1}^{\infty}$, where

$$D_i^{\mathcal{X}} \triangleq \Big( \mathcal{H}(\mathcal{X}) + \frac{H(Z_1|Z_i,\mathcal{X},\mathcal{Y})}{1 - p_\mathsf{r}} \Big)(1 - p_\mathsf{d}) - \frac{1 - p_\mathsf{d}}{1 - p_\mathsf{r}} h_2(p_\mathsf{r}) - h_2(p_\mathsf{d}), \ i \in \mathbb{N}$$

is non-decreasing.

*Proof:* Since we have introduced extra conditioning, the $D_i^{\mathcal{X}}$s are lower bounds. We have

$$\begin{aligned} H(Z_1|Z_{i+1}) &= H(Z_1, Z_i|Z_{i+1}) - H(Z_i|Z_1, Z_{i+1}) \\ &= H(Z_i|Z_{i+1}) + H(Z_1|Z_{[i:i+1]}) - H(Z_i|Z_1, Z_{i+1}) \\ &\overset{(a)}{=} H(Z_i|Z_{i+1}) + H(Z_1|Z_i) - H(Z_i|Z_1, Z_{i+1}) \\ &= H(Z_1|Z_i) + I(Z_1; Z_i|Z_{i+1}) \\ &\geq H(Z_1|Z_i) \end{aligned}$$

where $(a)$ follows from the Markovity of the $\mathcal{Z}$ process. Since conditioning on $\mathcal{X}$ and $\mathcal{Y}$ preserves the above chain of inequalities, we have $H(Z_1|Z_{i+1},\mathcal{X},\mathcal{Y}) \geq H(Z_1|Z_i,\mathcal{X},\mathcal{Y}) \ \forall \ i \geq 1$. Hence $\{D_i^{\mathcal{X}}\}_{i=1}^{\infty}$ is non-decreasing.

Optimizing $D_1^{\mathcal{X}}$ over stationary, ergodic, Markov input processes $\mathcal{X}$ gives the bound in Proposition 11. Therefore, for increasing $i$, we have bounds better than the one in Proposition 11. In particular, as $i \to \infty$, following the proof of Lemma 6 (iii), we can see that $\frac{Z_i}{i} \to \frac{p_\mathsf{r}-p_\mathsf{d}}{1-p_\mathsf{d}}$ a.s., so that the knowledge of $Z_i$ becomes tautological in the limit, and consequently,

$$\sup_{\mathcal{X}} \lim_{i\to\infty} D_i^{\mathcal{X}} = \sup_{\mathcal{X}} \sup_{i\geq 1} D_i^{\mathcal{X}}$$

gives us the bound in Proposition 14. $\blacksquare$

Alternatively, instead of bounding the information rate as in Proposition 14, we can write the following as an immediate consequence of Equation (10) and an argument similar to the one made in the proof of Proposition 14.

*Proposition 16 (Information rates for the DRC):* For the DRC,

$$\begin{aligned} C &= \sup_{\mathcal{X}} \Big( (1 - p_\mathsf{d})\mathcal{H}(\mathcal{X}) - \hat{\mathcal{H}}(\mathcal{Z}|\mathcal{Y}) + \hat{\mathcal{H}}(\mathcal{Z}|\mathcal{X},\mathcal{Y}) \Big) \\ &= (1 - p_\mathsf{d}) \Big[ \sup_{\mathcal{X}} \Big( \mathcal{H}(\mathcal{X}) + \frac{H(Z_1|\mathcal{X},\mathcal{Y}) - H(Z_1|\mathcal{Y})}{1 - p_\mathsf{r}} \Big) \Big]. \end{aligned} \qquad \blacksquare$$

$$\mathfrak{H}_m^{(i)} = \sum_{x_{[m+i-1]} \in \mathbb{X}^{m+i-1}} \frac{1}{2^{m+i-1}} \mathsf{H}(x_{[m+i-1]}),$$
$$\mathsf{H}(x_{[m+i-1]}) = \sum_{y_{[i-1]} \in \mathbb{Y}^{i-1}} \frac{w_{y_{[i-1]}}(x_{[m+i-1]})}{\binom{m+i-1}{m}} \mathfrak{h}(x_{[m+i-1]}, y_{[i-1]}), \tag{11}$$
$$\mathfrak{h}(x_{[m+i-1]}, y_{[i-1]}) = -\sum_{z=-m}^{0} \mathbb{1}_{\{x_{1-z}=y_1\}} \frac{w_{y_{[2:i-1]}}(x_{[2-z:m+i-1]})}{w_{y_{[i-1]}}(x_{[m+i-1]})} \log_2 \left( \mathbb{1}_{\{x_{1-z}=y_1\}} \frac{w_{y_{[2:i-1]}}(x_{[2-z:m+i-1]})}{w_{y_{[i-1]}}(x_{[m+i-1]})} \right).$$

Following arguments similar to the ones used in Lemma 15, we can show the following.

*Lemma 17:* The sequence of lower bounds $\{R_i^{\mathcal{X}}\}_{i=1}^{\infty}$, where

$$R_i^{\mathcal{X}} \triangleq (1-p_{\mathsf{d}})\Big(\mathcal{H}(\mathcal{X}) + \frac{H(Z_1|Z_i,\mathcal{X},\mathcal{Y}) - H(Z_1|\mathcal{Y})}{1-p_{\mathsf{r}}}\Big)$$

is non-decreasing, and

$$C = \sup_{\mathcal{X}} \lim_{i\to\infty} R_i^{\mathcal{X}} = \sup_{\mathcal{X}} \sup_{i\geq 1} R_i^{\mathcal{X}}. \qquad \blacksquare$$

The task of finding the rate-maximizing input distributions appears to be tough, with no theoretical insights[2] or efficient numerical algorithms. Often, to establish lower bounds on achievable rates, special classes of input processes are considered, and we will resort to a similar strategy here to obtain some expressions for the bounds we have so far developed. The following section will consider special cases of the DRC wherein there are either only deletions, i.e., the BDC, or only replications, i.e., the BRC. In a subsequent section, the symmetric DRC will be studied. All bounds developed in the next section are similar to the generic bounds developed thus far.

## IV. Channels with Deletions or Replications

For the case of the BDC or the BRC, evaluating some of the bounds developed in the previous section is somewhat easy, owing to the fact that the $\mathcal{Z}$ process is monotonic in these two special cases, i.e., it is non-increasing or non-decreasing with increments of at most one, respectively. This monotonicity in $\mathcal{Z}$ implies that the $\Gamma$ process is strictly increasing for the BDC and non-decreasing with increments of at most one for the BRC. This translates to the output being a subsequence of the input sequence for the BDC and vice versa for the BRC.

### A. Information Rates for the BDC

In this subsection we estimate the information rates possible over the BDC, i.e., $p_{\mathsf{d}} = p, p_{\mathsf{r}} = 0$, when the input process is either i.u.d. or when it is a first-order Markov process.

For the BDC with i.u.d. inputs, we can easily show that $\mathcal{Y}$ is also an i.u.d. sequence. Consequently,

$$\frac{I(Y_{[N_n]}; Z_{[N_n]})}{n} \to 0 \text{ as } n \to \infty$$

because the only information obtained from $Y_{[N_n]}$ about $Z_{[N_n]}$ is the length of the vector, and this information vanishes in the limit as $n \to \infty$. Therefore, we have from Equation (10) that the lower bound in Proposition 14 is actually the *symmetric information rate* (SIR). We are hence interested in evaluating $D_i^{\text{iud}}$ as defined in Lemma 15, where the superscript "iud" stands for independent, uniformly distributed inputs. In particular, we have the SIR

$$C_{\text{BDC}}^{\text{iud}} = \lim_{i\to\infty} D_i^{\text{iud}} = \sup_{i\geq 1} D_i^{\text{iud}}. \tag{12}$$

We start with some definitions and notation.

*Definition 18 (Subsequence weights):* We call a vector $x_A$ a *subsequence* of a vector $x_B$ if $A \subset B$ and the order of the elements in $A$ is the same as the order in which those elements appear in $B$. For ease of notation, we will write $w_{y_{[i]}}(x_{[j]})$ to denote the number of subsequences of $x_{[j]} \in \mathbb{X}^j$ that are the same as $y_{[i]} \in \mathbb{X}^i$, which is referred to as the $y_{[i]}$-*subsequence weight* of the vector $x_{[j]}$. We can write

$$w_{y_{[i]}}(x_{[j]}) = \sum_{S\subset[j]:|S|=i} \mathbb{1}_{\{x_S = y_{[i]}\}}$$

where the elements of the set $S$ are arranged in ascending order. Clearly, $w_{y_{[i]}}(x_{[j]}) = 0$ for $i > j$. We define $w_{\lambda}(x_{[j]}) = 1 \ \forall \ x_{[j]} \in \mathbb{X}^j$ for $j \geq 0$. $\square$

*Definition 19 (Runs and run-lengths):* For a binary sequence, a *run* is a maximal block of contiguous 0s or 1s. The *run-length* of a run is the number of symbols in it. We denote by $r_1(x_{[j]})$ the length of the first run in the vector $x_{[j]} \in \mathbb{X}^j, j \geq 1$. Clearly, $1 \leq r_1(x_{[j]}) \leq |x_{[j]}| = j$. $\square$

We will denote by $\mathbb{Z}_{\uparrow}^i$ and $\mathbb{Z}_{\downarrow}^i$ the sets of non-decreasing and non-increasing vectors of length $i$, respectively, for $i \geq 1$.

*Theorem 20 (SIR for the BDC):* For the BDC,

$$C_{\text{BDC}}^{\text{iud}} = 1 - p - h_2(p) + (1-p)\Big(\lim_{i\to\infty} \sum_{m\geq 0} \psi_{i,m} p^m (1-p)^i\Big),$$

where $\psi_{i,m} \triangleq \binom{m+i-1}{m}\mathfrak{H}_m^{(i)}$, with $\mathfrak{H}_m^{(i)} = H(Z_1|Z_i = -m, \mathcal{X}, \mathcal{Y})$ is as given in Equation (11).

*Proof:* For the BDC, we have from Lemma 15 that

$$D_i^{\text{iud}} = 1 - p - h_2(p) + (1-p)H(Z_1|Z_i,\mathcal{X},\mathcal{Y}).$$

From Equation (12), we need to show that

$$H(Z_1|Z_i,\mathcal{X},\mathcal{Y}) = \sum_{m\geq 0} \psi_{i,m} p^m (1-p)^i.$$

[2]A new result on BDC with small deletion probability [17] provides a partial answer to this question.

We first note that

$$H(Z_1|Z_i,\mathcal{X},\mathcal{Y}) = H(Z_1|Z_i, X_{[i-Z_i-1]}, Y_{[i-1]}).$$

Clearly, the above entropy term is zero for $i = 1$. For $i \geq 2$, given $Z_i = -m, X_{[m+i-1]} = x_{[m+i-1]}$ and $Y_{[i-1]} = y_{[i-1]}$, it is easy to see that

$$Z_1 \in \{z \in \{0,-1,\cdots,-m\} : x_{1-z} = y_1, \\ w_{y_{[2:i-1]}}(x_{[2-z:m+i-1]}) > 0\}.$$

That is, $Z_1 = z$ only if $x_{1-z}$ and $y_1$ match, and the subsequent part of the output vector $y_{[2:i-1]}$ is a subsequence of the subsequent part of the input vector $x_{[2-z:m+i-1]}$. Also, for $z_{[i-1]} \in \mathbb{Z}_{\downarrow}^{i-1}$ (which, as noted earlier, is true for the BDC),

$$\mathsf{P}(Z_{[i-1]} = z_{[i-1]}, Z_i = -m|X_{[m+i-1]} = x_{[m+i-1]}, \\ Y_{[i-1]} = y_{[i-1]}) = \mathbb{1}_{\{x_{[i-1]-z_{[i-1]}} = y_{[i-1]}\}} p_{\mathsf{t}}^i p_{\mathsf{d}}^m,$$

where $p_{\mathsf{d}} = 1 - p_{\mathsf{t}} = p$, so that for $0 \geq z \geq -m$,

$$\mathsf{P}(Z_1 = z, Z_i = -m|X_{[m+i-1]} = x_{[m+i-1]}, Y_{[i-1]} = y_{[i-1]}) \\ = \mathbb{1}_{\{x_{1-z}=y_1\}} p_{\mathsf{t}}^i p_{\mathsf{d}}^m \cdot w_{y_{[2:i-1]}}(x_{[2-z:m+i-1]}),$$

$$\mathsf{P}(Z_i = -m|X_{[m+i-1]} = x_{[m+i-1]}, Y_{[i-1]} = y_{[i-1]}) \\ = w_{y_{[i-1]}}(x_{[m+i-1]}) p_{\mathsf{t}}^i p_{\mathsf{d}}^m$$

and hence, when $w_{y_{[i-1]}}(x_{[m+i-1]}) > 0$,

$$\mathsf{P}(Z_1 = z|Z_i = -m, X_{[m+i-1]} = x_{[m+i-1]}, Y_{[i-1]} = y_{[i-1]}) \\ = \frac{\mathbb{1}_{\{x_{1-z}=y_1\}} w_{y_{[2:i-1]}}(x_{[2-z_1:m+i-1]}) p_{\mathsf{t}}^i p_{\mathsf{d}}^m}{w_{y_{[i-1]}}(x_{[m+i-1]}) p_{\mathsf{t}}^i p_{\mathsf{d}}^m} \\ = \frac{\mathbb{1}_{\{x_{1-z}=y_1\}} w_{y_{[2:i-1]}}(x_{[2-z_1:m+i-1]})}{w_{y_{[i-1]}}(x_{[m+i-1]})}.$$

Since, with i.u.d. inputs, $\mathsf{P}(X_{[m+i-1]} = x_{[m+i-1]}|Z_i = -m) = 2^{-(m+i-1)}$ and

$$\mathsf{P}(Y_{[i-1]} = y_{[i-1]}|X_{[m+i-1]} = x_{[m+i-1]}, Z_i = -m) \\ = \frac{w_{y_{[i-1]}}(x_{[m+i-1]})}{\binom{m+i-1}{m}},$$

we have that $H(Z_1|Z_i = -m, \mathcal{X}, \mathcal{Y}) = \mathfrak{H}_m^{(i)}$ as in Equation (11). By noting that

$$\mathsf{P}(Z_i = -m|Z_0 = 0) = \binom{m+i-1}{m} p^m (1-p)^i$$

from Equation (7) (with $p_{\mathsf{d}} = p, p_{\mathsf{r}} = 0, p_{\mathsf{t}} = 1 - p$), we have the desired result. ∎

Although evaluating $\mathfrak{H}_m^{(i)}$ in general is hard since we are required to count subsequence weights of sequences, we can evaluate it in two specific cases: for every $m$ when $i = 2$ (when all but a single bit are deleted) and for all $i$ when $m = 1$ (when only a single bit is deleted). We examine these two cases in detail in Appendix V-A, V-B and state the results here.

*Corollary 21 (Lower bound for $C_{\text{BDC}}^{\text{iud}}$):* For the BDC,

$$C_{\text{BDC}}^{\text{iud}} \geq D_2^{\text{iud}} \geq \frac{4(1-p)^3}{(2-p)^2} - h_2(p) \\ + (1-p)^3 \Big( \sum_{m \geq 2} m p^{m-1} \log_2 m \Big). \quad ∎$$

*Corollary 22 (Small deletion probability SIR):* For the BDC,

$$C_{\text{BDC}}^{\text{iud}} = 1 + p \log_2 p - \mathtt{d} p + O(p^2)$$

where $\mathtt{d} \approx 1.154163765$. ∎

Similar bounds for symmetric first-order Markov input processes are considered in Appendix V-C. Fig. 1 plots the bounds for $C_{\text{BDC}}$.

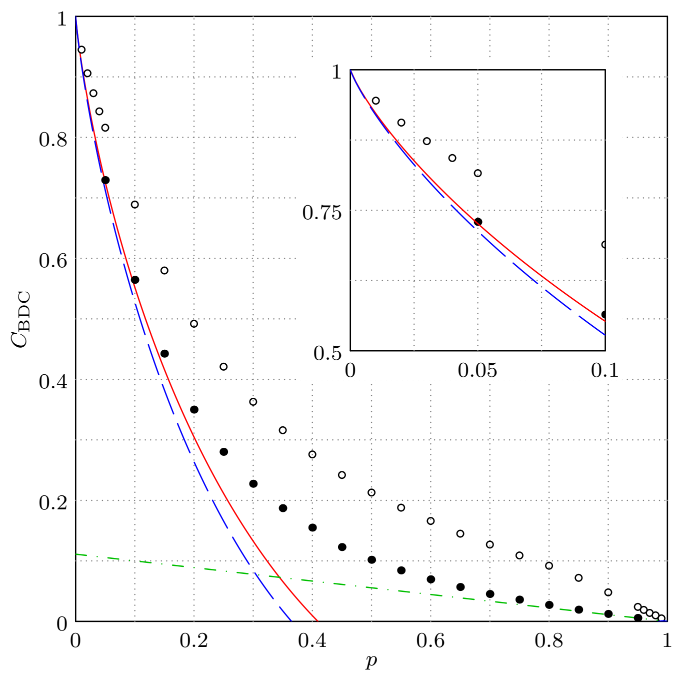


Fig. 1. Bounds on the capacity for the BDC in bits per channel use as a function of the deletion probability $p$. $D_2^{\text{iud}}$ (cf. Equation (16)) is shown as the long-dashed blue line and $C^{\text{iud}}$ (or equivalently $\mathfrak{D}_1^{\text{iud}}$) with the $O(p^2)$ term dropped as the solid red line (cf. Equation (19)). The best known numerical lower [11] and upper bounds [15] are shown as black and white circles respectively. The best known lower bound as $p$ approaches 1 [9] is shown as the dash-dotted green line. The inset shows the bounds for small $p$ values where the red solid curve is known to be tight from [12].

## B. Information Rates for the BRC

In this subsection, we will consider information rates for the BRC, i.e., $p_{\mathsf{d}} = 0, p_{\mathsf{r}} = p$. As in the previous subsection, we will consider i.u.d. and symmetric first-order Markov inputs.

For the BRC, the $\mathcal{Z}$ process is non-decreasing. Moreover, when it increases, the increment is at most 1 at each time instant. This simplifies the evaluation of information rates and we will, in fact, be able to write exact expressions for the Markov-1 rates, as will be shown shortly. In this case, even when the input is i.u.d., the term

$$\frac{I(Y_{[N_n]}; Z_{[N_n]})}{n} \nrightarrow 0 \text{ as } n \to \infty$$

in the normalized version of Equation (10). Hence the expression for the information rate in Proposition 16 will prove to be more useful in this subsection.

*Theorem 23 (Markov-1 Rates for the BRC):* For the BRC, the Markov-1 rate is given as in Equation (13).

$$C_{\mathrm{BRC}}^{\mathcal{M}1} = \max_{\alpha} \Big[ h_2(\alpha) + \alpha \sum_{l\geq 1} \Big( (1-\alpha)\frac{1-p}{p}\Big)^l \Big( \sum_{k\geq l} \binom{k}{l} p^k h_2(\frac{l}{k})\Big) - \frac{p+(1-\alpha)(1-p)}{1-p} h_2\Big(\frac{p}{p+(1-\alpha)(1-p)}\Big)\Big]. \quad (13)$$

*Proof:* First we note that since $Z_1 \in \{0,1\}$, we have $H(Z_1|\mathcal{X},\mathcal{Y}) = \mathsf{E}[h_2(\mathsf{P}(Z_1 = 0 \mid x_{\mathbb{N}},\ y_{\mathbb{N}}))]$ and $H(Z_1 \mid \mathcal{Y}) = \mathsf{E}[h_2(\mathsf{P}(Z_1 = 0 \mid y_{\mathbb{N}}))]$. Further we have for the BRC that whenever $Y_i \neq Y_{i-1}$, we must have $Z_i = Z_{i-1}$ or equivalently $\Gamma_i = \Gamma_{i-1}+1$. This means that $Z_1$ is independent of subsequent runs of $\mathcal{Y}$ (and $\mathcal{X}$) given the first run of $\mathcal{Y}$ (and $\mathcal{X}$) since we can achieve synchronization at the end of each run. Thus we can write the conditional probabilities $\mathsf{P}(Z_1 = 0 \mid \mathcal{X},\ \mathcal{Y})$ and $\mathsf{P}(Z_1 = 0 \mid \mathcal{Y})$ in terms of the first runs of $\mathcal{X}$ and $\mathcal{Y}$, i.e., $\mathsf{P}(Z_1 = 0 \mid x_{\mathbb{N}},\ y_{\mathbb{N}}) = \mathsf{P}(Z_1 = 0 \mid r_1(x_{\mathbb{N}}),\ r_1(y_{\mathbb{N}}))$ and $\mathsf{P}(Z_1 = 0 \mid y_{\mathbb{N}}) = \mathsf{P}(Z_1 = 0 \mid r_1(y_{\mathbb{N}}))$. Note that we assume that $Z_0 = 0$ so that $Y_0 = X_0$. Thus, if $x_1 \neq x_0$, then $Z_1$ is 0 or 1 accordingly as $y_1$ is not equal or equal to $y_0$, respectively. This means that there is no uncertainty in $Z_1$ given the output sequence (and the assumption that $x_0 = y_0 = 0$, which can be made without loss of generality). Therefore, in estimating the entropy of $Z_1$ given the output sequence, or the output and the input sequences, we can confine our attention to those sequences $x_{\mathbb{N}}$ and $y_{\mathbb{N}}$ whose first runs are comprised of zeros. We shall denote such runs as $r_1^0(\cdot)$. For a first-order Markov input process, we have, for $l \geq 0$

$$\mathsf{P}(r_1^0(x_{\mathbb{N}}) = l) = (1-\alpha)^l \alpha,$$

and we can get from the definition of the BRC that

$$\mathsf{P}(r_1^0(y_{\mathbb{N}}) = k | r_1^0(x_{\mathbb{N}}) = l) = \binom{k}{l}(1-p)^{l+1}p^{k-l}$$

for $k \geq l$. Consequently, we have

$$\begin{aligned}\mathsf{P}(r_1^0(y_{\mathbb{N}}) = k) &= \sum_{l=0}^{k}(1-\alpha)^l\alpha \cdot \binom{k}{l}(1-p)^{l+1}p^{k-l} \\ &= \alpha(1-p)\Big(p+(1-\alpha)(1-p)\Big)^k.\end{aligned}$$

Since $Z_1 = 0$ excludes the first bit in the received sequence from being a replication, we can easily obtain

$$\mathsf{P}(Z_1 = 0 | r_1^0(x_{\mathbb{N}}) = l, r_1^0(y_{\mathbb{N}}) = k) = \frac{l}{k}$$

for $k \geq l + \mathbb{1}_{\{l=0\}}$. For $k \geq 1$,

$$\begin{aligned}&\mathsf{P}(Z_1 = 0 | r_1^0(y_{\mathbb{N}}) = k) \\ &\quad = \frac{\sum_{l=0}^k \mathsf{P}(Z_1 = 0, r_1^0(x_{\mathbb{N}}) = l, r_1^0(y_{\mathbb{N}}) = k)}{\mathsf{P}(r_1^0(y_{\mathbb{N}}) = k)} \\ &\quad = \frac{\sum_{l=0}^k (1-\alpha)^l \alpha \binom{k}{l}(1-p)^{l+1}p^{k-l}(\frac{l}{k})}{(1-p)\alpha\Big(p+(1-\alpha)(1-p)\Big)^k} \\ &\quad = \frac{(1-\alpha)(1-p)}{p+(1-\alpha)(1-p)}.\end{aligned}$$

Therefore,

$$\begin{aligned}H(Z_1|\mathcal{X},\mathcal{Y}) &= \sum_{l\geq 0}(1-\alpha)^l\alpha \\ &\quad \times \Big(\sum_{k\geq l+\mathbb{1}_{\{l=0\}}} \binom{k}{l}(1-p)^{l+1}p^{k-l}h_2(\frac{l}{k})\Big) \\ &= \alpha(1-p)\sum_{l\geq 1}\Big((1-\alpha)\frac{1-p}{p}\Big)^l\Big(\sum_{k\geq l}\binom{k}{l}p^k h_2(\frac{l}{k})\Big)\end{aligned}$$

and

$$\begin{aligned}H(Z_1|\mathcal{Y}) &= \sum_{k\geq 1}\alpha(1-p)\Big(p+(1-\alpha)(1-p)\Big)^k \\ &\qquad \times h_2\Big(\frac{(1-\alpha)(1-p)}{p+(1-\alpha)(1-p)}\Big) \\ &= (p+(1-\alpha)(1-p))h_2\Big(\frac{p}{p+(1-\alpha)(1-p)}\Big).\end{aligned}$$

Substituting these in Proposition 16 specialized to the BRC and first-order Markov inputs, we have the desired result. ■

The following results are shown in Appendix VI.

*Corollary 24 (Lower bound for $C_{\mathrm{BRC}}^{\mathcal{M}1}$):* For the BRC,

$$\begin{aligned}C_{\mathrm{BRC}}^{\mathcal{M}1} \geq R_2^{\mathcal{M}1} = h_2&\Big(\frac{1}{(1-p)(4^p+1)}\Big) \\ &+ \Big(\frac{2p}{1-p}\Big)\Big(\frac{(1-p)4^p - p}{4^p+1}\Big) \\ &- \Big(\frac{1}{1-p}\Big)\Big(\frac{4^p}{4^p+1}\Big)h_2\Big(\frac{p(4^p+1)}{4^p}\Big)\end{aligned}$$

for $0 \leq p \leq p_* \approx 0.734675821$. ■

*Corollary 25 (Small replication probability SIR):* For the BRC,

$$C_{\mathrm{BRC}}^{\mathrm{iud}} = 1 + p\log_2 p + \mathtt{r}p + O(p^2)$$

where $\mathtt{r} \approx 0.845836235$. ■

Fig. 2 plots these bounds. Note that the SIR and the Markov-1 rate are non-convex in $p$. Further, it appears that the Markov-1 rate (and the SIR) are zero for some values of $p < 1$. However, this behavior is due to the fact that the term

$$\sum_{k\geq l}\binom{k}{l}p^k h_2\Big(\frac{l}{k}\Big)$$

in Equation (13) is computed only up to a finite value of $k$ (the curves in Fig. 2 are, therefore, lower bounds for the Markov-1 rate and the SIR). For values of $p$ close to 1, more terms in this sum need to be considered to get a better estimate of the achievable rates.

*Remark 2:* It was expected that the capacity of a memoryless SEC was a convex function of the channel parameters. Although this conjecture seems to be true for the BDC [19],

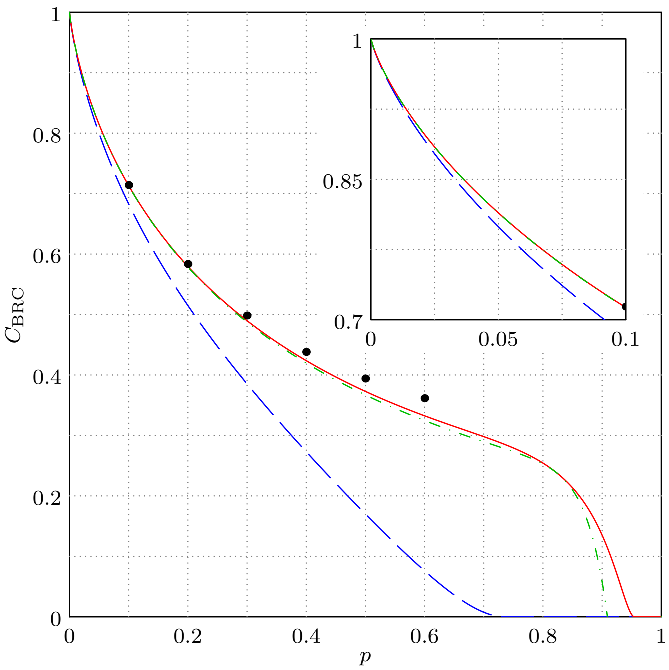


Fig. 2. Lower bounds on the capacity for the BRC. The bound $R_2^{\mathcal{M}1}$ from Corollary 24 is shown as the long-dashed blue line and the Markov-1 rate in Equation (13) is shown as the solid red line. The SIR ($\alpha = \frac{1}{2}$ in Equation(13)) is the dash-dotted green line. The numerical lower bounds in [5] are shown as black circles. The inset shows the bounds for small replication probabilities.

we see that this conjecture is false for the BRC. Note that the lower bounds in [5] themselves lead one to question the conjecture (cf. Fig. 2). However, the Markov-1 rate for the BRC in Equation (13) settles this conjecture as being false for the BRC. This is because if the capacity were convex in the replication probability, no rate larger than $(1-p)$ would be achievable, which is clearly not the case as can be seen from Fig. 2. This implies that, in general, in presence of synchronization errors, the capacity is not convex in the channel parameters. In particular, it is possible that the capacity for the BDC is non-convex as well.

## V. Channels with Deletions and Replications

Although the bounds in the previous section provide us some idea of the achievable information rates for the BDC and the BRC, they do not generalize in a straightforward manner for an SEC with both deletions and replications[3]. In order to obtain bounds when both deletions and replications are present, we take a different approach.

### A. Approximate Non-Stationary Channels

We construct a sequence of channels that approximate the DRC $\mathbf{P}_n$. To this end, we fix $m \in \mathbb{Z}^+$ and let

$$Z_i^{(m)} = \begin{cases} Z_i, & |Z_i| \le m \\ m \cdot \mathrm{sgn}(Z_i), & |Z_i| > m \end{cases} \tag{14}$$

[3]It is possible to obtain, albeit with a lot more effort than in the cases of the BDC or the BRC, the lower bound $D_2^{\mathcal{X}}$ for a first-order Markov input process for a DRC. We shall omit this here.

for $Z_i$s given by the $\mathcal{Z}$ process defined in Section III with $\mathrm{sgn} : \mathbb{R} \mapsto \{\pm 1\}$ defined as

$$\mathrm{sgn}(x) = \begin{cases} 1, & x \ge 0 \\ -1, & x < 0. \end{cases}$$

We then define the channel model for the $m^{\text{th}}$-approximating channel $\mathbf{P}_{n,m}^{\dagger} = (\mathbb{X}, \mathbb{Y}, P_{n,m}^{\dagger})$ as was done for $\mathbf{P}_n$,

$$Y_i^{(m)} = X_{\Gamma_i^{(m)}} = X_{i - Z_i^{(m)}}, i \in [N_n^{(m)}]$$

where $N_n^{(m)} = \sup\{i \ge 0 : \Gamma_i^{(m)} \le n | \Gamma_0^{(m)} = 0\}$. It is clear that $n - m \le N_n^{(m)} \le n + m$.

The input and output alphabets of the channel $\mathbf{P}_{n,m}^{\dagger}$ are $\mathbb{X}$ and $\mathbb{Y}$ respectively, same as those of $\mathbf{P}_n$. The transition probability $P_{n,m}^{\dagger}$ for the channel $\mathbf{P}_{n,m}^{\dagger}$ is defined as in Equation (5), but with the channel states defined by the process $\mathcal{Z}^{(m)} \triangleq \{Z_i^{(m)}\}_{i \ge 1}$. The transition probability of the $\mathcal{Z}^{(m)}$ process itself is defined by Equations (14) and (7).

We now establish a few properties of the $\mathcal{Z}^{(m)}$ process and the approximating channels $\mathbf{P}_{n,m}^{\dagger}$. We start with some properties of the state process $\mathcal{Z}^{(m)}$ and the index process $\Gamma^{(m)}$ that will be useful in proving subsequent results. The following property, proved in Appendix VII, establishes the non-stationarity of the sequence of channels $\{\mathbf{P}_{n,m}^{\dagger}\}_{m \ge 0}$.

*Lemma 26 (Properties of $\mathcal{Z}^{(m)}$):* The state process $\mathcal{Z}^{(m)}$ is a finite, time-inhomogeneous Markov chain. Moreover, the boundary states $\{\pm m\}$ are eventually *absorbing* states, under the measure $\mathsf{P}$, in the following two cases.

(i) For $m = o(n)$ when $p_{\mathsf{d}} \ne p_{\mathsf{r}}$.
(ii) For $m = o(\sqrt{n})$ when $p_{\mathsf{d}} = p_{\mathsf{r}}$. ■

*Lemma 27 (Refinedness of $\{\Gamma^{(m)}\}_{m \ge 0}$):* For a fixed $n \in \mathbb{N}$, for every $m \in \mathbb{Z}^+$,

$$\{\Gamma_{[N_n]}\} \cap [n] \subset \{\Gamma^{(m)}_{[N_n^{(m)}]}\} \cap [n] \text{ a.s.}$$

and

$$\{\Gamma^{(m+1)}_{[N_n^{(m+1)}]}\} \cap [n] \subset \{\Gamma^{(m)}_{[N_n^{(m)}]}\} \cap [n] \text{ a.s.,}$$

where $\{\Gamma_{\mathbb{U}}\}$ denotes the set of elements of the random vector $\Gamma_{\mathbb{U}}$, i.e., where the random variables are not repeated. ■

The proof is given in Appendix VIII.

*Proposition 28:* For every $m \in \mathbb{Z}^+$,

$$I(X_{[n]}; Y_{[N_n]}) \le I(X_{[n]}; Y^{(m)}_{[N_n^{(m)}]}), \text{ and}$$
$$I(X_{[n]}; Y^{(m+1)}_{[N_n^{(m+1)}]}) \le I(X_{[n]}; Y^{(m)}_{[N_n^{(m)}]}). \quad ■$$

The proof is left to Appendix IX. Intuitively, the above result is true because the "drift" between the input and the output processes is bounded by $m$ for the approximating channel $\mathbf{P}_{n,m}^{\dagger}$, whereas it is unbounded for the DRC $\mathbf{P}_n$ (or equivalently $\mathbf{Q}_n$). The result below, which gives a total ordering of the sequence of channels $\{\mathbf{P}_{n,m}^{\dagger}\}_{m \ge 0}$ in terms of their mutual information rates, follows immediately from Proposition 28.

*Corollary 29 (Total Ordering of $\{\mathbf{P}^{\dagger}_{n,m}\}_{m\geq 0}$):* For any $n \in \mathbb{N}$, the sequence $\{I^{\dagger}_{n,m}\}_{m\geq 0}$, where

$$I^{\dagger}_{n,m} \triangleq \frac{1}{n} I(X_{[n]}; Y^{(m)}_{[N^{(m)}_n]}),$$

is non-increasing. Since $I^{\dagger}_{n,m} \in [0,1]\ \forall\ n \in \mathbb{N}$ and $m \in \mathbb{Z}^+$, $\lim_{m\to\infty} I^{\dagger}_{n,m}$ exists and is equal to $\inf_{m\geq 0} I^{\dagger}_{n,m}$. ∎

*Proposition 30 (Information limits):* For any $n \in \mathbb{N}$, we have

$$I_n \triangleq \frac{1}{n} I(X_{[n]}; Y_{[N_n]}) = I^{\dagger}_n \triangleq \lim_{m\to\infty} \downarrow I^{\dagger}_{n,m} = \inf_{m\geq 0} I^{\dagger}_{n,m}.$$

Consequently, for a stationary, ergodic input process $\mathcal{X}$,

$$I_{\mathcal{X}} \triangleq \lim_{n\to\infty} I_n = \inf_{n\geq 1} I_n = I^{\dagger}_{\mathcal{X}} \triangleq \lim_{n\to\infty} I^{\dagger}_n = \inf_{n\geq 1} I^{\dagger}_n,$$

so that

$$C = \sup_{\mathcal{X}} I_{\mathcal{X}} = \sup_{\mathcal{X}} I^{\dagger}_{\mathcal{X}}$$

for stationary, ergodic, Markov processes $\mathcal{X}$.

*Proof:* The last equality in the first line is from Corollary 29. From Proposition 28, we have $I_n \leq I^{\dagger}_{n,m}\ \forall\ m \in \mathbb{Z}^+$, from which $I_n \leq I^{\dagger}_n$ follows. The equality is true because of the following. If we let $\mathscr{F}_{n,m} \triangleq \sigma(\{X_{[n]}, Y^{(m)}_{[N^{(m)}_n]}\})$, the sigma-algebra generated by the random variables $\{X_{[n]}, Y^{(m)}_{[N^{(m)}_n]}\}$, then $\{\mathscr{F}_{n,m}\}_{m\geq 0}$ is a *filtration* [20, §10.1], i.e.,

$$\mathscr{F}_{n,m} \subset \mathscr{F}_{n,m+1}\ \forall\ m \geq 0.$$

Thus, $\mathsf{P}(X_{[n]}, Y^{(m)}_{[N^{(m)}_n]})$ is the restriction of $\mathsf{P}$ to $\mathscr{F}_{n,m}$. From [21, Theorem 2], we have that $I_n = I^{\dagger}_n$.

The limit of $I_n$ as $n$ goes to infinity exists and is equal to the infimum of the sequence from the subadditivity of the sequence $\{nI_n\}_{n\geq 1}$ and Fekete's Lemma (cf. [22, Appendix II]). The last claim made is true from Proposition 3. ∎

*Corollary 31:* For any $n \in \mathbb{N}$, we have that

$$C_n = \sup_{\mathsf{P}(X_{[n]})} I_n = C^{\dagger}_n \triangleq \sup_{\mathsf{P}(X_{[n]})} I^{\dagger}_n$$

where $C_n$ is as defined in Theorem 2. Therefore

$$C = \lim_{n\to\infty} C_n = \inf_{n\geq 1} C_n = C^{\dagger} \triangleq \lim_{n\to\infty} C^{\dagger}_n.$$ ∎

Although $\{\mathbf{P}^{\dagger}_{n,m}\}_{m\geq 0}$ is a sequence of channels that approximate $\mathbf{P}_n$, and have the properties discussed so far in this subsection, they are not useful as finite-state channels (FSCs), as shown below.

*Lemma 32 (FSCs $\mathbf{P}^{\dagger}_{n,m}$):* For any $m \in \mathbb{Z}^+$, for a stationary, ergodic input process $\mathcal{X}$,

$$I^{\dagger}_{\mathcal{X}}(m) \triangleq \lim_{n\to\infty} I^{\dagger}_{n,m} = \mathcal{H}(\mathcal{X})$$

so that

$$C^{\dagger}(m) \triangleq \sup_{\mathcal{X}} I^{\dagger}_{\mathcal{X}}(m) = 1.$$

*Proof:* From Lemma 26, the states $\{\pm m\}$ are eventually absorbing for any $m \in \mathbb{Z}^+$. Hence, in the limit as $n \to \infty$, the channel only has a delay of $\pm m$, and hence the result. ∎

We now attempt to obtain approximate channels that are stationary and useful as FSCs.

### B. Approximate Stationary Channels

Let $m \in \mathbb{Z}^+$. Fix $n \in \mathbb{N}$. Consider the channel $\mathbf{P}^{\star}_{n,m} = (\mathbb{X}, \mathbb{Y}, P^{\star}_{n,m})$ where

$$Y^{(m)}_i = X_{\Gamma^{(m)}_i} = X_{i - Z^{(m)}_i}, i \in [N^{(m)}_n]$$

with $N^{(m)}_n, \Gamma^{(m)}_{[N^{(m)}_n]}$ and $Z^{(m)}_{[N^{(m)}_n]}$ as defined for the channel $\mathbf{P}^{\dagger}_{n,m}$. The difference will be in the underlying measure $\mathsf{P}_{\langle m\rangle}$. Let the measure $\mathsf{P}_{\langle m\rangle}$ be such that the $\mathcal{Z}^{(m)}$ process is a finite, time-homogeneous, first-order Markov chain with transition probabilities

$$\mathsf{P}_{\langle m\rangle}(Z^{(m)}_i = k | Z^{(m)}_{i-1} = j) = \mathsf{P}(Z^{(m)}_i = k | Z^{(m)}_{i-1} = j)$$

when $-m < j < m$,

$$\mathsf{P}_{\langle m\rangle}(Z^{(m)}_i = k | Z^{(m)}_{i-1} = -m) = \begin{cases} 1 - p_{\mathsf{r}}, & k = -m \\ p_{\mathsf{r}}, & k = -m+1 \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\mathsf{P}_{\langle m\rangle}(Z^{(m)}_i = k | Z^{(m)}_{i-1} = m) = \begin{cases} 1 - p_{\mathsf{d}}(1-p_{\mathsf{r}}), & k = m \\ (1-p_{\mathsf{d}})(1-p_{\mathsf{r}})p_{\mathsf{d}}^{m-k}, & -m < k < m \\ (1-p_{\mathsf{r}})p_{\mathsf{d}}^{2m}, & k = -m \\ 0, & \text{otherwise}. \end{cases}$$

Note that the measure $\mathsf{P}_{\langle m\rangle}$ differs from $\mathsf{P}$ only for state paths that reach beyond the states $\{\pm m\}$. The transition probabilities $P^{\star}_{n,m}$ for the channel $\mathbf{P}^{\star}_{n,m}$ can now be defined as in Equation (5), but under the measure $\mathsf{P}_{\langle m\rangle}$. The stationarity of the channels $\mathbf{P}^{\star}_{n,m}$ follows from the time-homogeneity of the $\mathcal{Z}^{(m)}$ process.

*Remark 3:* Note that the sequence of sigma-algebras $\{\mathscr{G}_m\}_{m\geq 0}$ where $\mathscr{G}_m \triangleq \sigma(\mathcal{Z}^{(m)})$ forms a filtration. The sequence of measures $\mathsf{P}_{\langle m\rangle}$ as defined above seem to be defined only on the corresponding sigma-algebras $\mathscr{G}_m$s for each $m \in \mathbb{Z}^+$. However, we can extend these measures to the sigma-algebra $\mathscr{B}$ as in Appendix X, and will henceforth consider $\mathsf{P}_{\langle m\rangle} : \mathscr{B} \mapsto [0,1]$ for each $m \in \mathbb{Z}^+$. □

The lemma below shows that for a fixed $m \in \mathbb{Z}^+$, the FSC $\mathbf{P}^{\star}_{n,m}$ is an indecomposable FSC [23, §4.6].

*Lemma 33 ($\mathbf{P}^{\star}_{n,m}$ Indecomposable):* The FSC $\mathbf{P}^{\star}_{n,m}$ is indecomposable for every $m \in \mathbb{Z}^+$ for $(p_{\mathsf{d}}, p_{\mathsf{r}}) \in (0,1)^2$.

*Proof:* Fix $m \in \mathbb{Z}^+$. We need to make a couple of modifications to put the channels $\{\mathbf{P}^{\star}_{n,m}\}_{n\geq 1}$ in the parlance of discrete FSCs. First, we set

$$\acute{Y}^{(m)}_i = Y^{(m)}_{i-m} = X_{i-m-Z^{(m)}_{i-m}} = X_{i-\acute{Z}^{(m)}_i} \text{ for } i \in [n].$$

Note that $\acute{Z}_i^{(m)} = m + Z_{i-m}^{(m)} \in [0:2m]$, and hence the channel producing $\acute{Y}_{[n]}^{(m)}$ is "causal". Let the "state" $W_i^{(m)}$ of the channel $\mathbf{P}^\star_{n,m}$ at time $i \in [n]$ be defined as

$$W_i^{(m)} = (X_{[i-2m:i-1]}, \acute{Z}_i^{(m)}) \in \mathbb{X}^{2m} \times [0:2m],$$

where we set $X_i = 0$ for $i \notin [n]$. Note that we need to redefine the state of the channel in this case to keep the factorization

$$\begin{aligned}\mathsf{P}_{\langle m\rangle}(\acute{Y}_i^{(m)}, W_{i+1}^{(m)}|X_i, W_i^{(m)}) \\ = \mathsf{P}_{\langle m\rangle}(\acute{Y}_i^{(m)}|X_i, W_i^{(m)}) \cdot \mathsf{P}_{\langle m\rangle}(W_{i+1}^{(m)}|X_i, W_i^{(m)}).\end{aligned}$$

Since $\acute{\mathcal{Z}}^{(m)}$ is a finitely delayed, finitely shifted version of $\mathcal{Z}^{(m)}$, and because $\mathcal{Z}^{(m)}$ is an *irreducible, aperiodic* Markov chain under the measure $\mathsf{P}_{\langle m\rangle}$ [24, Chapter 1] as long as $(p_\mathsf{d}, p_\mathsf{r}) \in (0,1)^2$, so is $\acute{\mathcal{Z}}^{(m)}$. In particular, we have that for every $i \geq 2m$,

$$\min_{z\in[0:2m]} \mathsf{P}_{\langle m\rangle}(\acute{Z}_i^{(m)} = z|\acute{Z}_0^{(m)} = z') > 0 \;\forall\; z' \in [0:2m].$$

This implies that for $i = 2m$ and $\overline{x} \in \overline{\mathbb{X}}$, by choosing[4] $w = (x_{[0:2m-1]}, z)$ for any $z \in [0:2m]$, we see that

$$\mathsf{P}_{\langle m\rangle}(W_{2m}^{(m)} = w|\overline{X} = \overline{x}, W_0^{(m)} = w') > 0$$

for every $w' \in \mathbb{X}^{2m} \times [0:2m]$. From [23, Theorem 4.6.3], we have the desired result. ∎

*Remark 4:* Note that in the description of the causal channel in the proof above, we have discarded the part of the output $Y^{(m)}_{[n-m+1:N_n^{(m)}]}$ by considering the causal output $\acute{Y}_{[n]}^{(m)}$. This will however not matter in the estimation of the information rate since

$$0 \leq \frac{1}{n} I(X_{[n]}; Y^{(m)}_{[N_n^{(m)}]}) - \frac{1}{n} I(X_{[n]}; \acute{Y}^{(m)}_{[n]}) \leq \frac{m}{n},$$

and since $m \in \mathbb{Z}^+$ is fixed, the rates are the same in the limit as $n$ goes to infinity. □

*Corollary 34 (Capacity of FSC $\mathbf{P}^\star_{n,m}$):* For $m \in \mathbb{Z}^+$ and the sequence of FSCs $\{\mathbf{P}^\star_{n,m}\}_{n\geq 1}$ with $(p_\mathsf{d}, p_\mathsf{r}) \in (0,1)^2$, the capacity is given by

$$\begin{aligned} C^\star(m) &= \lim_{n\to\infty} \sup_{\mathsf{P}_{\langle m\rangle}(X_{[n]})} \frac{1}{n} I(X_{[n]}; Y^{(m)}_{[N_n^{(m)}]}) \\ &\triangleq \lim_{n\to\infty} \sup_{\mathsf{P}_{\langle m\rangle}(X_{[n]})} I^\star_{n,m}.\end{aligned}$$

*Proof:* From Lemma 33 and Remark 4, we have $C^\star(m)$ as defined in the statement can be written as

$$C^\star(m) = \lim_{n\to\infty} \sup_{\mathsf{P}_{\langle m\rangle}(X_{[n]})} \frac{1}{n} I(X_{[n]}; \acute{Y}^{(m)}_{[n]}).$$

Now since

$$\begin{aligned} I(X_{[n]}; \acute{Y}^{(m)}_{[n]}) &= I(X_{[n]}; W_0^{(m)}) + I(X_{[n]}; \acute{Y}^{(m)}_{[n]}|W_0^{(m)}) \\ &\quad - I(X_{[n]}; W_0^{(m)}|\acute{Y}^{(m)}_{[n]}),\end{aligned}$$

we have

$$\begin{aligned} |I(X_{[n]}; \acute{Y}^{(m)}_{[n]}) - I(X_{[n]};\acute{Y}^{(m)}_{[n]}|W_0^{(m)})| \\ \leq \log_2\left((2m+1)|\mathbb{X}|^{2m}\right) \\ = 2m\log_2|\mathbb{X}| + \log_2(2m+1).\end{aligned}$$

Therefore

$$C^\star(m) = \lim_{n\to\infty} \sup_{\mathsf{P}_{\langle m\rangle}(X_{[n]})} \frac{1}{n} I(X_{[n]}; \acute{Y}^{(m)}_{[n]}|W_0^{(m)}).$$

From [23, Theorem 4.6.4], the quantity on the right hand side of the above equality exists and is the capacity of the indecomposable FSCs $\{\mathbf{P}^\star_{n,m}\}_{n\geq 1}$. ∎

*Corollary 35 (Capacity of $\mathbf{P}^\star_{n,m}$):* For the FSCs $\{\mathbf{P}^\star_{n,m}\}_{n\geq 1}$, the capacity $C^\star(m)$ can be written [22] as

$$\begin{aligned} C^\star(m) &= \sup_{\mathcal{X}} \lim_{n\to\infty} \frac{1}{n} I(X_{[n]}; Y^{(m)}_{[N_n^{(m)}]}) \\ &= \sup_{\mathcal{X}} \lim_{n\to\infty} I^\star_{n,m} \triangleq \sup_{\mathcal{X}} I^\star_{\mathcal{X}}(m)\end{aligned}$$

where the supremum is over all stationary, ergodic input sources $\mathcal{X}$. ∎

From Lemma 33, since $\{\mathbf{P}^\star_{n,m}\}_{n\geq 1}$ are indecomposable FSCs, we have from [25] that

$$\begin{aligned} -\frac{1}{n}\log_2 \mathsf{P}_{\langle m\rangle}(X_{[n]}, Y^{(m)}_{[N_n^{(m)}]}) &\to \lim_{n\to\infty} \frac{H(X_{[n]}, Y^{(m)}_{[N_n^{(m)}]})}{n} \\ &= \hat{\mathcal{H}}(\mathcal{X}, \mathcal{Y}^{(m)}), \\ -\frac{1}{n}\log_2 \mathsf{P}_{\langle m\rangle}(Y^{(m)}_{[N_n^{(m)}]}) &\to \lim_{n\to\infty} \frac{H(Y^{(m)}_{[N_n^{(m)}]})}{n} \\ &= \hat{\mathcal{H}}(\mathcal{Y}^{(m)}),\end{aligned}$$

as $n \to \infty$ a.s., where the entropies are calculated with respect to the measure $\mathsf{P}_{\langle m\rangle}$. Therefore

$$I^\star_{\mathcal{X}}(m) = \mathcal{H}(\mathcal{X}) + \hat{\mathcal{H}}(\mathcal{Y}^{(m)}) - \hat{\mathcal{H}}(\mathcal{X}, \mathcal{Y}^{(m)})$$

can be estimated numerically using the forward passes of the BCJR algorithm [26] to estimate $\hat{\mathcal{H}}(\mathcal{X}, \mathcal{Y}^{(m)})$ and $\hat{\mathcal{H}}(\mathcal{Y}^{(m)})$, as in [27], [28]. Moreover, optimizing Markov input sources numerically is possible [29], [30] for these FSCs.

In Fig. 3, we plot the SIRs, $C^\star_{\text{iud}}(m)$, for the indecomposable FSCs $\{\mathbf{P}^\star_{n,m}\}_{n\geq 1}$ obtained through numerical simulations for $1 \leq m \leq 8$ and $p_\mathsf{d} = p_\mathsf{r} = p \in [0, \frac{1}{2}]$. The value of $n$ used for the estimation was $5 \times 10^5$. The error in estimation is consequently upper bounded by $0.15\%$.

A couple of observations are worthwhile noting. First, the SIRs $\{C^\star_{\text{iud}}(m)\}_{m\geq 0}$ are non-increasing. This hints at a total ordering of the FSCs $\{\mathbf{P}^\star_{n,m}\}_{m\geq 0}$ with respect to the information rates similar to what we had in Corollary 29. Second, we see that for small values of $p$, the SIRs get bunched up as $m$ increases, i.e., the SIRs $C^\star_{\text{iud}}(m)$ converge quickly, so that we have a good estimate of

$$C^\star_{\text{iud}}(\infty) \triangleq \lim_{m\to\infty} C^\star_{\text{iud}}(m)$$

for $p$ close to $0$.

[4]Set $x_0 = 0$ by convention.

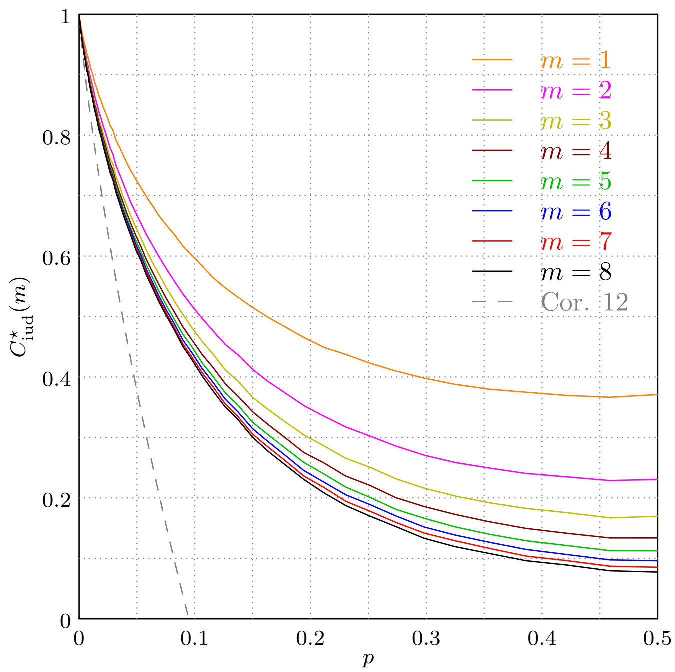


Fig. 3. SIR estimates for the FSCs $\{\mathbf{P}^\star_{n,m}\}_{n\geq 1}$ with $p_{\mathsf{d}} = p_{\mathsf{r}} = p \in [0, \frac{1}{2}]$ for different $m$ values are shown in solid lines. The lower bound on the capacity of the SDRC from Corollary 12 is also shown as the dashed line.

*Proposition 36:* For $n \in \mathbb{N}$, we have

$$I^\star_n \triangleq \liminf_{m\to\infty} I^\star_{n,m} = I_n.$$

Thus,

$$C = \sup_{\mathcal{X}} \inf_{n\geq 1} \liminf_{m\to\infty} I^\star_{n,m}.$$

*Proof:* For a fixed $n \in \mathbb{N}$, since we have that $\mathsf{P}_{\langle m\rangle}(X_{[n]}, Y^{(m)}_{[N^{(m)}_n]}) \to \mathsf{P}(X_{[n]}, Y_{[N_n]})$ as $m \to \infty$ for every $\vartheta \in \mathbb{S}$, $\mathsf{P}_{\langle m\rangle}(X_{[n]}, Y^{(m)}_{[N^{(m)}_n]})$ converges to $\mathsf{P}(X_{[n]}, Y_{[N_n]})$ in total variation as $m \to \infty$. Consequently, from [31, Corollary $1'$], we have the desired result. ∎

*Remark 5:* We conjecture that $I^\star_{n,m} \to I^\star_n = I_n$ as $m \to \infty$. From [31, Corollary $1'$], one needs to show uniform integrability of the *information densities* $i(X_{[n]}, Y^{(m)}_{[N^{(m)}_n]})$ for the conjecture to be true. Alternatively, if the sequence of channels $\{\mathbf{P}^\star_{n,m}\}_{m\geq 0}$ is totally ordered for every $n \in \mathbb{N}$ with respect to the mutual information rates, as was the case for the sequence $\{\mathbf{P}^\dagger_{n,m}\}_{m\geq 0}$ (cf. Corollary 29), i.e., if $\{I^\star_{n,m}\}_{m\geq 0}$ is a non-increasing sequence for every $n \in \mathbb{N}$, then we know that

$$\lim_{m\to\infty} I^\star_{n,m} = I^\star_n,$$

and from Proposition 36, $I^\star_{n,m} \downarrow I_n$ follows. Unfortunately, we are not able to show this monotonicity in the sequence $\{I^\star_{n,m}\}_{m\geq 0}$ as we argued in the case of the sequence $\{I^\dagger_{n,m}\}_{m\geq 0}$. Although the refinedness property of the $\Gamma$ process (cf. Lemma 27) still holds, the different measures $\mathsf{P}_{\langle m\rangle}$ being used for each $m \in \mathbb{Z}^+$ do not allow us to generalize the result of Corollary 29. However, Fig. 3 provides sufficient empirical evidence for this monotonicity conjecture. □

### *C. Approximating Channels for the SDRC*

In this subsection, we consider the SDRC, i.e., the case when $p_{\mathsf{d}} = p_{\mathsf{r}} = p \in [0, 1)$. This channel is of interest since in practice, systems prone to mis-synchronization are usually not biased to produce more deletions or replications. For the case of the SDRC, we can fix $m$ to be a function of $n$ satisfying a simple condition and define a sequence of approximating channels.

*Lemma 37:* For the SDRC, for every $n \in \mathbb{N}$, let $m \in \mathbb{N}$. Then,

$$\begin{aligned}&\mathsf{P}_{\langle m\rangle}\Big(\max_{i=1}^{N^{(m)}_n} |Z_i| \geq m\Big)\\ &\qquad = \mathsf{P}\Big(\max_{i=1}^{N^{(m)}_n} |Z_i| \geq m\Big) = O\Big(\frac{n+m}{m^2}\Big).\end{aligned}$$ ∎

We relegate the proof to Appendix XI. The significance of the above result can be seen by noticing that, for the SDRC, the probability (under measure $\mathsf{P}$ or $\mathsf{P}_{\langle m\rangle}$) with which the approximating channels introduced in the previous two subsections differ from the actual channel can be made arbitrarily small by setting $m(n) = \omega(\sqrt{n})$, i.e., $\lim_{n\to\infty} \frac{m(n)}{\sqrt{n}} = \infty$, and choosing a large enough $n$. For the so-chosen sequence of approximating channels, we can conclude that the limiting channel characterizes the SDRC from the following result whose proof is left to Appendix XII.

*Proposition 38 (Approximating SDRC):* For the SDRC,

$$I_{\mathcal{X}} = \lim_{n\to\infty} I_n = \liminf_{n\to\infty} I^\star_{n,m(n)}$$

where $m(n) = \omega(\sqrt{n})$, for stationary, ergodic input process $\mathcal{X}$. ∎

The channels $\{\mathbf{P}^\star_{n,m}\}_{m\geq 0}$ give us a way to approach the problems of optimizing input distributions as well as designing coding schemes for the SDRC. We can optimize the inputs of $\mathbf{P}^\star_{n,m}$, starting with small values of $m$, under some input assumptions, e.g., for fixed-order Markov inputs [29], [30]. Note that the numerical estimation of $I^\star_{n,m}$ is possible (as described in the previous subsection) only when $m < n$, since setting the channels as indecomposable FSCs (cf. Lemma 33) is possible only in this case. Moreover, for a good estimate of the information rate, we will require $m \ll n$. For the SDRC, Proposition 38 allows us to consider some $\mathbf{P}^\star_{n,m(n)}$, where $m(n)$ is both $\omega(\sqrt{n})$ as well as $o(n)$, for which a good estimate of the information rate $I^\star_{n,m(n)}$ can be obtained. Note that due to the lack of a result analogous to Lemma 37 in the case of a general DRC for $m < n$, generalizing these arguments when $p_{\mathsf{d}} \neq p_{\mathsf{r}}$ is not completely justified.

Starting with some small values of $m$, we expect that the information rates and optimal distributions quickly converge (in $m$), giving us a way to characterize optimal inputs for the SDRC $\mathbf{P}_n$. For small values of $p$, as in Fig. 3, the information rates for the SDRC can be characterized numerically for moderate values of $m$ (much smaller than $\omega(\sqrt{n})$ guaranteed

by Lemma 37). For optimizing the input distribution for an approximation $\mathbf{P}^{\star}_{n,m}$, we can start with optimizing inputs that are $\mu^{\text{th}}$-order Markov processes, for $\mu \geq 1$. As was observed[5] in [32], the convergence of optimal information rates as a function of the order $\mu$ of the input Markov process is expected to be rapid. The authors in [32] hypothesized that this convergence was exponential in $\mu$. Similar "diminishing returns" on increasing $\mu$ has also been observed by others [29], [30]. We think that a similar rapid convergence of $I^{\star}_{n,m}(\mathcal{X}^{*}_{\mathcal{M}\mu})$ to $C_n(\mathcal{X}^{*}_{\mathcal{M}\mu})$ also holds for $m$, where $I^{\star}_{n,m}(\mathcal{X}^{*}_{\mathcal{M}\mu})$ is the optimal information rate achieved by a $\mu^{\text{th}}$-order Markov input process on the FSC $\mathbf{P}^{\star}_{n,m}$ and $C_n(\mathcal{X}^{*}_{\mathcal{M}\mu})$ is the optimal information rate achieved by a $\mu^{\text{th}}$-order Markov input process on the SDRC $\mathbf{P}_n$. We use the *generalized Blahut-Arimoto algorithm* presented in [30] to evaluate $I^{\star}_{n,m}(\mathcal{X}^{*}_{\mathcal{M}\mu})$ for some small values of $m$ and $\mu$. Figure 4 plots these estimates, which illustrates the aformentioned observations. Note that it

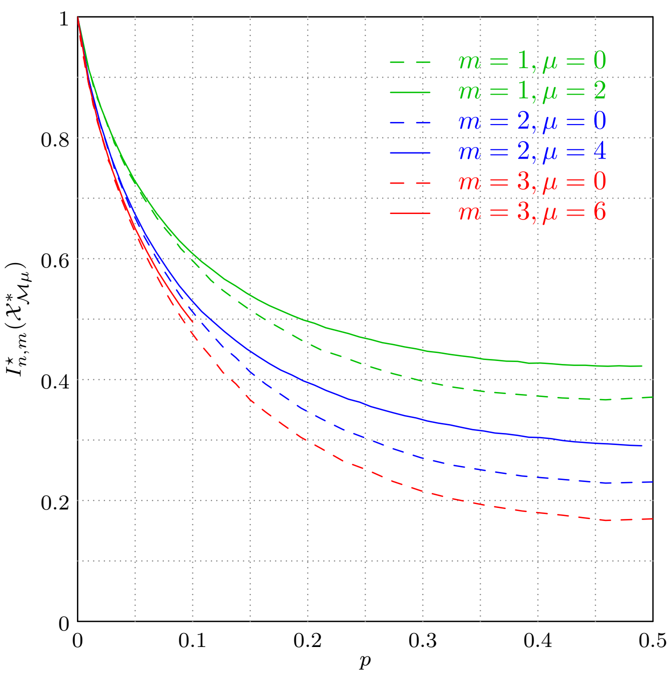


Fig. 4. Numerical estimates of $I^{\star}_{n,m}(\mathcal{X}^{*}_{\mathcal{M}\mu})$ for $m = 1, 2$ and 3, and $\mu = 2m$ (solid lines). For the case where $m = 3, \mu = 6$, only the estimates for small deletion-replication probabilities have been evaluated. We chose $n = 10^6$ for $m = 1, 2$ and $n = 10^5$ for $m = 3$. The smaller value of $n$ in the case of $m = 3$ was chosen for computational convenience. Also shown (in dashed lines) for comparison are the corresponding estimates of the SIR ($\mu = 0$) from Figure 3.

is clear from the plots above that Bernoulli equiprobable inputs achieve rates (SIR) comparable to higher-order Markov inputs for small proabablities of deletion and replication. However, unlike ISI channels [33], it is not clear that Markov sources of increasing orders achieve capacity. It is also evident from the figure that for small values of the deletion-replication probabilities, the information rates seem to converge as $m$ increases even for small values of $m$ ($m = 3$). This suggests that Figure

[5]Although the validity of the bounds in [32] is unclear (See, e.g., [10]), the rapid convergence of information rates as a function of the order $\mu$ of the input Markov process is expected to be true.

4 plots good estimates of $I^{\star}_{n,m(n)}$ for such channels, and since the values of $n$ chosen are large, that they indeed represent good estimates of the Markov capacity of the SDRC.

Apart from the above advantage of facilitating numerical estimation of information rates, the approximating channels $\mathbf{P}^{\star}_{n,m}$ have another important advantage. This is that since they have immediate factor-graph interpretations, there is a possibility of constructing sparse graph-based coding schemes and decoding over the joint graphical model representing the channels as well as the codes, as was done for joint detection and decoding of LDPC codes on partial response channels [34]. Instead of trying to build codes for the SDRC $\mathbf{P}_n$, the problem can be reduced to designing good codes and efficient decoding schemes for the FSCs $\mathbf{P}^{\star}_{n,m}$ for small values of $m$. For small deletion-replication probabilities $p$, which is the case of interest in practice, we can expect these codes to perform well for the SDRC $\mathbf{P}_n$ as well.

## VI. Generalizations

In this section, we discuss different scenarios that can be modeled using the channel model introduced in Section III with appropriate modifications. Wherever possible, methods for information theoretic analysis for these cases through generalizations of the channel model presented in this paper are highlighted.

### A. *Channels introducing Random Insertions*

The channel model of Equation (6) allows us to handle deletions as well as replications. However, the class of SECs that introduce random insertions cannot be written in the form of Equation (6). A suitable modification for our model in this scenario is to let

$$Y_i = X_{i-Z_i} \oplus \mathbb{1}_{\{Z_i = Z_{i-1}+1\}} V_i,$$

where $\mathbb{X} = \mathbb{Y} = \mathbb{V} = \{0, 1\}$, and $\mathcal{V} = \{V_i\}_{i \geq 1}$ is a Bernoulli sequence with parameter $f$ (for "flip"). This means that the probability of a random insertion is $f p_{\mathsf{r}}$ and that of a replication is $(1-f)p_{\mathsf{r}}$. Note that this can be easily generalized to any finite sets $\mathbb{X}$, and arbitrary sets $\mathbb{Y}$ and $\mathbb{V}$ (with an appropriate notion of the addition operation "$\oplus$"). Analysis of this channel is, however, more complicated than analyzing the DRC itself due to the cascaded additive noise channel which also depends on the "shared" state process $\mathcal{Z}$. However, when the channel produces deletions, replications, random insertions as well as substitutions, we can write

$$Y_i = X_{i-Z_i} \oplus V_i,$$

which is just a cascade of the DRC and an additive noise channel. In the binary setting, this corresponds to a channel that deletes a bit $x$ with probability $p_{\mathsf{d}}$, or inserts a sequence $\overline{y}$ with probability $(1 - p_{\mathsf{d}})(1 - p_{\mathsf{r}})p_{\mathsf{r}}^{|\overline{y}|-1} f^{w(\overline{y}+x)}$. This implies that the substitution error probability for a bit is given by $(1 - p_{\mathsf{d}})(1 - p_{\mathsf{r}})f$.

The capacity (or the information rate achievable by a given input process) and coding for a cascade of binary, memoryless channels without synchronization errors has been studied in, e.g., [35]–[38]. Some lower bounds on the capacity of a

cascade of a BDC and an additive noise channel were given in [16], [39]. The possibility of extending these results to general memoryless SECs using the model presented here remains a problem worth exploring.

### B. SECs with Memory

An SEC with memory is defined as in Definition 1, with the only difference being that the transition probabilities $q_n(\overline{y}_{[n]}|x_{[n]})$ do not factorize as products of individual probabilities $q(\overline{y}_i|x_i)$. As an example, one could think of an SEC where having a deletion for a symbol influences the likelihood of the following symbol being deleted, i.e., a channel that introduces a burst of deletions. Similarly, channels that introduce a limited number of deletions in every input subsequence of a certain length could also occur in practice. These have been studied under the name of *segmented* deletion channels [40], [41]. Note that the definition of the segmented deletion channel is slightly different in the references cited, where it is assumed that the input is divided into blocks of a certain length and at most one deletion occurs within each block. Our definition is more general and corresponds closer to reality.

The channel model in Equation (6) generalizes readily to the case of DRC with memory. Consider the $\mathcal{Z}$ process to be a non-increasing (so that only deletions occur), time-homogeneous, shift-invariant Markov process of order $\mathsf{z} \geq 2$ such that

$$\mathsf{P}(Z_i = z_i | Z_{[i-\mathsf{z}:i-1]} = z_{[i-\mathsf{z}:i-1]}) > 0$$

only for $z_{[i-\mathsf{z}:i-1]}$ such that $z_{i-\mathsf{z}} = z_{i-\mathsf{z}+1} = \cdots = z_{i-j} \geq z_{i-j+1} = \cdots = z_{i-1}$ for some $1 \leq j \leq \mathsf{z}$, and $z_{i-j} - z_{i-j+1} \leq 1$. Then, clearly at most one deletion occurs for every input subsequence of length $\mathsf{z}$. The model in Equation (6) will then correspond to a segmented deletion channel where no more than 1 deletion occurs for every $\mathsf{z}$ input symbols. Similarly, we can model other DRC with memory by suitably considering the $\mathcal{Z}$ process to be a Markov process of some order with specific transitions occuring with non-zero probability.

Although we have let $\mathsf{z} \geq 2$, not all second-order Markov processes $\mathcal{Z}$ result in SECs with memory. One example worth noting is when the $\mathcal{Z}$ process is non-decreasing with increments of at most 1, and is such that two consecutive increments occur with probability 0. This results in a replication channel where each symbol is transmitted noiselessly and possibly replicated once—this is exactly the *elementary sticky channel* introduced in [5], which is a memoryless SEC. We will refer to such channels that introduce a bounded number of inserted symbols per input symbol as *bounded*, memoryless SECs. This particular channel has also been studied in [16], where some analytical lower bounds on the capacity were given. Another example where $\mathsf{z} = 2$ does not result in an SEC with memory but is a bounded, memoryless SEC is Gallager's model [7] of the insertion-deletion channel. Some analytical lower bounds for the capacity of this channel (without deletions) were given in [16]. Achievable rates for a bounded, memoryless SEC were studied in [42], and those for a cascade of a bounded, memoryless SEC with an inter-symbol interference (ISI) channel in [43]. Some bounds on the capacity of a bounded, memoryless SEC with substitution errors were given in [44].

Note that the channel coding theorem for SECs with memory has not been established. The various works on the "capacity" of such channels is an indication of such SECs occurring widely in practice. Establishing the channel coding theorem for SECs with memory is, therefore, important both for the theory and in practice. For SECs with memory, since the channel model (6) will have the transition probabilities that still factorize as in Equation (5) (with the channel state transition probabilities replaced by the higher-order transition probabilities), it is more amenable to analysis and could potentially be used to establish the channel coding theorem.

### C. Jitter, Bit-shift and Grain-error Channels

Channels that consider mis-synchronization due to *jitter* or *bit-shifts* have been studied in the context of magnetic recording and constrained coding [45]–[47]. These represent a variant of the general model of the DRC presented in the present paper. In particular, they are characterized by a $\mathcal{Z}$ process where each valid state path $\overline{z} \in \overline{\mathbb{Z}}$ has increments and decrements of size at most 1, and the transition probabilities are data-dependent. The zeros and ones in the input correspond to the absence and presence, respectively, of a transition in the signal. Thus, the presence of a transition cannot be deleted, i.e., a 1 in the input stream cannot be deleted, whereas the 0s can be deleted or replicated (at most once). The authors of [45] gave bounds on the capacity and the zero-error capacity of bit-shift channels and also present some bounds on achievable rates over a concatenation of the bit-shift channel with a binary symmetric channel. Similar analysis was performed in [46] for discrete and continuous channels with timing jitter. Numerical upper and lower bounds on the capacity of a binary channel with jitter where transitions could "cancel" each other were given in [47].

Another class of channels that resemble these channels are the "paired" insertion-deletion channels studied in the context of bit-patterned media recording [48]. Here, the channel is similar to the approximating FSC given in Section V-B with $m = 1$. In [48], the authors give bounds on the capacity and the zero-error capacity of the channel for varying sizes of the state space. A further specialization of this channel is the one-dimensional *graunlar media* recording channel. This has also been studied in [49], where some bounds on capacity and coding constructions have been proposed.

### D. Permuting and Trapdoor Channels

The *trapdoor channel* introduced by Blackwell (See [50, §7.1]) is a channel where the input stream is fed to a buffer at the same rate as symbols from within the buffer are randomly drawn as the output stream. Using our model, we can define the trapdoor channel as follows. The multiset of indices of the buffer contents at time $i \geq 1$ is denoted as $B_i = \{\beta_1, \cdots, \beta_b\}$, which is of size $b$. We initialize

$$B_1 = \{\underbrace{0, \cdots, 0}_{b-1}, 1\}$$

and define the output at the $i^{\text{th}}$ instant as $Y_i = X_{\Gamma_i}$ for $i \in [n]$, where $\Gamma_i$ has the distribution $\mathsf{P}(\Gamma_i = \beta_j) = \frac{1}{b}, 1 \le j \le b$. The buffer multiset is updated as $B_{i+1} = B_i \setminus \{\Gamma_i\} \cup \{i+1\}$. In this case, a further simplification of the channel model might be more useful since the channel depends not on the indices of the inputs in the buffer, but on the *type* [18, §12.1] of the buffer contents at any time. This channel was generalized to define *permuting channels* in [51].

Although the trapdoor channel is described easily, its capacity, even in the simplest case of $|\mathbb{X}| = 2$ and $b = 2$, has been an open problem ever since its introduction. In [52], the authors considered coding schemes for certain non-probabilistic models of the trapdoor channel. The capacity of the probabilistic trapdoor channel when $|\mathbb{X}| = 2$ and $b = 2$ is known to satisfy [53]

$$\frac{1}{2} < C(|\mathbb{X}| = 2, b = 2) \le \log_2 \frac{1+\sqrt{5}}{2} \approx 0.694241914.$$

It is worthwhile to explore the possibility of obtaining better bounds on the capacity of the trapdoor and permuting channels using the model presented in this paper.

### *E. Molecular Communication and Chemical Channels*

A simple model for molecular communication or chemical channels is as follows. The channel state $Z_i$ at time instant $i$ is a random variable on the alphabet $\{0\} \cup [m]$ and represents the delay introduced to the input at time $i$. The output at time $i$ is given as

$$Y_i = \sum_{\mathsf{z}=0}^{m} X_{i-\mathsf{z}} \mathbb{1}_{\{Z_{i-\mathsf{z}}=\mathsf{z}\}},$$

i.e., the output is the sum of all the channel inputs that arrive at time $i$. This channel was studied in [54] as a *delay selector channel* and a lower bound on the capacity was given assuming that the state process is i.i.d.. In general, the state process can be modeled as a Markov process, and the channel might be amenable to a similar analysis as presented here.

### *F. Timing Channels*

There is a link between discrete *timing channels* [55], where information is communicated not only in the signals but also in the timing of the signals and the randomness is in the arrival times of the signals, and "good" transmission sequences for SECs. This is in the sense that a information-bearing transmission sequence for an SEC must not only be able to carry information within the sequence, but also contain information in the ordering of the symbols within the sequence, such that even in the presence of synchronization errors, the information about the symbol ordering is not completely lost. That is to say that the sequences $X_{[n]}$ must be such that under limited number of synchronization errors, the received sequence $Y_{[N_n]}$ must convey adequate information about the state sequence $Z_{[N_n]}$. Therefore, it might be of importance to study whether methods of coding over timing can be used to obtain efficient codes and decoding schemes for the SECs.

## VII. Conclusions

We introduced a new channel model for a class of SECs which formulated the SEC as a channel with states. This allowed us to obtain analytical lower bounds for the capacity of SECs with only deletions or only replications. For the case of the BDC, we were able to write the SIR in terms of subsequence weights of binary sequences. Subsequence weights are known to be a quantity of interest in the maximum-likelihood decoding of sequences for the BDC (cf. Equation (11)). Moreover, it is clear from Equation (11) that the dependence of information rates for the BDC on the input statistics only appears in the term $\mathfrak{H}_m^{(i)}$, whereas the subsequence weights influence $\mathsf{H}(\overline{x})$ independently of the input statistics. Thus, our result establishes a natural link between the capacity of the BDC and the metric relevant for ML decoding. We were also able to obtain lower bounds on the capacity of the BDC that are known to be tight for small deletion probabilities. For the BRC, we were able to exactly characterize the Markov-1 rate, which is, to the best of our knowledge, the first analytic lower bound on the capacity of the BRC. In doing so, we were able to disprove the conjecture that the capacity of SECs is a convex function of the channel parameters, at least in the case of the BRC.

For the case of an SEC with deletions and replications, we were able to provide a sequence of approximating FSCs that are totally ordered with respect to the mutual information rates achievable, and therefore, with respect to capacities. These approximating FSCs were shown to be such that the mutual information rate achievable for the SEC was equal to the limit of the mutual information rates achievable for the sequence of FSCs. To obtain numerical estimates of achievable rates on the DRC, we defined another sequence of indecomposable FSCs. Computing the mutual information rates for this sequence of FSCs allows us to relate the mutual information rate for the DRC to the limiting value of the mutual information rates of the sequence. For the particular case of the SDRC, we were able to show a stronger form of convergence of these mutual information rates.

The formulation in this paper not only allows us to get estimates of mutual information rates achievable on SECs but also gives some insight into possible code constructions and decoding schemes for such channels. The approximations introduced for the DRC gives us a natural way to reduce these problems. One would therefore obtain progressively better performing codes for the DRC by designing good codes for the sequence of approximating FSCs. We expect that for a small values of the deletion-replication probability, a code constructed for an approximation with a moderate value of $m$ will perform well over the DRC as well. Some coding schemes for special cases of the FSCs (with $m = 1$) have been known in various contexts (See Section VI-C). Extending these schemes to better approximations (larger $m$ values) will prove crucial in designing good codes for the DRC. We emphasize that although the present paper considers only *binary* SECs, the results extend naturally to the case of larger finite alphabets. The expressions for information rates will perhaps become more complicated, but the methods to arrive at their bounds

or numerical estimates remain unchanged.

The present formulation of the SECs also allows us to make the following remarks on the BDC.

- In [17], the authors conjectured that the capacity of the BDC has a Taylor-like series expansion. We see from Theorem 20 that this is true for the SIR of the BDC. We expect that the capacity also has a similar formulation.
- The capacity of a general SEC might not be convex in the channel parameters (See Remark 2). It was shown in [19] that the capacity $C_{\mathrm{BDC}}$ of the BDC satisfies

$$\inf_{p\in[0,1]} \frac{C_{\mathrm{BDC}}(p)}{1-p} = \lim_{p\to 1} \frac{C_{\mathrm{BDC}}(p)}{1-p}.$$

  It is therefore expected that $C_{\mathrm{BDC}}(p)$ is convex in $p$. From Theorem 20, we see that the SIR $C_{\mathrm{BDC}}^{\mathrm{iud}}$ of the BDC can be written as

$$\begin{aligned} C_{\mathrm{BDC}}^{\mathrm{iud}} &= 1-p-h_2(p) \\ &\quad + (1-p)\Big(\lim_{i\to\infty}\lim_{v\to\infty}\sum_{m=0}^{v} f(i,m,p)\Big), \end{aligned}$$

  where $f(i,m,p) = \psi_{i,m}p^m(1-p)^i$ is non-convex in $p$ for $m \geq 1$. It is interesting to see if this double limit turns out to be convex despite

$$\sum_{m=0}^{v} f(i,m,p)$$

  being non-convex for every $v \geq 1$. Extending this to the case of the capacity $C_{\mathrm{BDC}}$ is also of interest.
- In order to obtain bounds for the capacity of a BDC for $p$ close to 1, one might typically consider the case where all but one (or a few) symbols are lost. The lower bound $D_2$ presented here (cf. Lemma 15) corresponds to this situation. However, since we considered this bound for a first-order Markov input, the bounds we obtained didn't prove to be useful for $p$ close to 1. It might therefore be of interest to generalize this bound for a high-order Markov input which might give us a strictly positive (and thereby non-trivial) achievable rate.

## Appendix I
## Proof of Lemma 6

(i) This is true since $p_{\mathsf{r}} < 1$.

(ii) Since $p_{\mathsf{d}} < 1$.

(iii) Notice that, for each $n \in \mathbb{N}$, we can write

$$\Gamma_n = n - Z_n = \sum_{i=1}^{n} \Delta_i$$

where the $\Delta_i$'s are i.i.d. with

$$\mathsf{P}(\Delta_1 = \delta) = \begin{cases} p_{\mathsf{r}}, & \delta = 0 \\ p_{\mathsf{t}}p_{\mathsf{d}}^{\delta-1}, & \delta \geq 1. \end{cases}$$

From the strong law of large numbers (SLLN), we therefore have $\frac{\Gamma_n}{n} \to \mathsf{E}(\Delta_1)$ a.s. as $n \to \infty$. We also have $N_n \to \infty$ a.s. as $n \to \infty$ from point (ii) above. Therefore, $\frac{\Gamma_{N_n}}{N_n} \to \mathsf{E}(\Delta_1)$ a.s. as $n \to \infty$. Further, by definition, we have $\Gamma_{N_n} \leq n < \Gamma_{N_n+1}$, i.e.,

$$\frac{\Gamma_{N_n}}{N_n} \leq \frac{n}{N_n} \leq \frac{\Gamma_{N_n+1}}{N_n+1}\Big(\frac{N_n+1}{N_n}\Big).$$

Thus $\frac{N_n}{n} \to \frac{1}{\mathsf{E}(\Delta_1)} = \frac{1-p_{\mathsf{d}}}{1-p_{\mathsf{r}}}$ a.s. as $n \to \infty$.

## Appendix II
## Proof of Lemma 7

(i) By definition, $\mathcal{Z}$ is a first-order Markov chain. Time-homogeneity implies that

$$\mathsf{P}(Z_i|Z_{i-1}) = \mathsf{P}(Z_1|Z_0)\ \forall\ i \geq 1.$$

This is true for the state process $\mathcal{Z}$ from the definition since the transition probabilities in Equation (7) do not depend on the time index $i$. Shift-invariance implies

$$\mathsf{P}(Z_1 = z_1|Z_0 = z_0) = \mathsf{P}(Z_1 = z_1 - z_0|Z_0 = 0).$$

This is true because the state transition probabilities in (7) depend only on the difference $z_i - z_{i-1}$.

The $\mathit{\Gamma}$ process inherits these properties from $\mathcal{Z}$ through the bijection $\zeta : \mathbb{Z}^n \mapsto \mathbb{Z}^n$, where with some abuse of notation, we write $\Gamma_{[n]} = \zeta(Z_{[n]}) = (\zeta(Z_i)), i \in [n]$, with $\Gamma_i = \zeta(Z_i) = i - Z_i, i \in [n],\ \forall\ n \in \mathbb{N}$.

The irreducibility and aperiodicity of the $\mathcal{Z}$ process follow from the definition.

(ii) Note that from Equation (7), $Z_{i+1} \leq Z_i + 1$ a.s. for every $i \geq 0$. Hence $Z_{i+j} \leq Z_i + j$ a.s. for every $i \geq 0, j \geq 0$. Since $\Gamma_i = i - Z_i$, we have $\Gamma_{i+j} = i + j - Z_{i+j} \geq i + j - Z_i - j = \Gamma_i$ with probability 1.

(iii) From the bijection $\zeta$ (See point (i) above) and Appendix I, we have

$$\begin{aligned} H(Z_i|Z_{i-1}) &= H(\Gamma_i|\Gamma_{i-1}) = H(\Gamma_{i-1} + \Phi_i|\Gamma_{i-1}) \\ &= H(\Phi_i|\Gamma_{i-1}) = H(\Phi_i) = H(\Phi_1) \\ &= h_2(p_{\mathsf{r}}) + \frac{1-p_{\mathsf{r}}}{1-p_{\mathsf{d}}}h_2(p_{\mathsf{d}}). \end{aligned}$$

Hence

$$\begin{aligned} H(Z_{[n]}) &= \sum_{i=1}^{n} H(Z_i|Z_{[i-1]}) = \sum_{i=1}^{n} H(Z_i|Z_{i-1}) \\ &= n\Big(h_2(p_{\mathsf{r}}) + \frac{1-p_{\mathsf{r}}}{1-p_{\mathsf{d}}}h_2(p_{\mathsf{d}})\Big). \end{aligned}$$

## APPENDIX III
## PROOF OF PROPOSITION 10

Let $Z_0 = 0$ and consider semi-infinite input, state and output processes. We first note that $\forall\ k, l \in \mathbb{N}, k \leq l$,

$$\begin{aligned}
&\mathsf{P}(Y_{[k:l]} = y_{[k:l]} | Z_{k-1} = 0)\\
&= \sum_{\gamma_{[k:l]}} \mathsf{P}(\Gamma_{[k:l]} = \gamma_{[k:l]}, X_{\gamma_{[k:l]}} = y_{[k:l]} | \Gamma_{k-1} = k-1)\\
&\overset{(a)}{=} \sum_{\gamma_{[k:l]}} \mathsf{P}(\Gamma_{[k:l]} = \gamma_{[k:l]} - z, X_{\gamma_{[k:l]}} = y_{[k:l]} |\\
&\qquad\qquad \Gamma_{k-1} = k-1-z)\\
&\overset{(b)}{=} \sum_{\gamma_{[k:l]}} \mathsf{P}(\Gamma_{[k:l]} = \gamma_{[k:l]} - z, X_{\gamma_{[k:l]}-z} = y_{[k:l]} |\\
&\qquad\qquad \Gamma_{k-1} = k-1-z)\\
&= \mathsf{P}(Y_{[k:l]} = y_{[k:l]} | \Gamma_{k-1} = k-1-z)\\
&= \mathsf{P}(Y_{[k:l]} = y_{[k:l]} | Z_{k-1} = z)\ \forall\ z \leq k-1.
\end{aligned}$$

Here, $(a)$ follows from the shift-invariance of $\mathit{\Gamma}$ (See Lemma 7) and $(b)$ from the stationarity of $\mathcal{X}$. Therefore, we have

$$\begin{aligned}
\mathsf{P}(Y_{[k]} = y_{[k]}) &= \sum_{z\in\mathbb{Z}} \mathsf{P}(Z_0 = z)\mathsf{P}(Y_{[k]} = y_{[k]} | Z_0 = z)\\
&= \mathsf{P}(Y_{[k]} = y_{[k]} | Z_0 = 0)\\
&= \mathsf{P}(Y_{[k]} = y_{[k]} | \Gamma_0 = 0)\\
&= \sum_{\gamma_{[k]}} \mathsf{P}(\Gamma_{[k]} = \gamma_{[k]}, X_{\gamma_{[k]}} = y_{[k]} | \Gamma_0 = 0)\\
&\overset{(c)}{=} \sum_{\gamma_{[k]}} \mathsf{P}(\Gamma_{[j+1:j+k]} = \gamma_{[k]}, X_{\gamma_{[k]}} = y_{[k]} | \Gamma_j = 0)\\
&= \mathsf{P}(Y_{[j+1:j+k]} = y_{[k]} | Z_j = 0)\\
&= \mathsf{P}(Y_{[j+1:j+k]} = y_{[k]})\ \forall\ j, k \in \mathbb{N}
\end{aligned}$$

where $(c)$ follows form the time-homogeneity of $\mathit{\Gamma}$ (Lemma 7). The last equality above follows from the observation made in the beginning of the proof.

## APPENDIX IV
## PROOF OF LEMMA 13

From Equation (8), we can write

$$\begin{aligned}
(i+j)H_{i+j} &= H(Z_{[N_{i+j}]} | X_{[i+j]}, Y_{[N_{i+j}]})\\
&= H(Z_{[N_i]} | X_{[i+j]}, Y_{[N_{i+j}]})\\
&\quad + H(Z_{[N_i+1:N_{i+j}]} | X_{[i+j]}, Y_{[N_{i+j}]}, Z_{[N_i]})\\
&\geq H(Z_{[N_i]} | X_{[i+j]}, Y_{[N_{i+j}]}, N_i)\\
&\quad + H(Z_{[N_i+1:N_{i+j}]} | X_{[i+j]}, Y_{[N_{i+j}]}, Z_{[N_i]}, N_i)\\
&\overset{(a)}{=} H(Z_{[N_i]} | X_{[i]}, Y_{[N_i]})\\
&\quad + H(Z_{[N_i+1:N_{i+j}]} | X_{[i+1:i+j]}, Y_{[N_i+1:N_{i+j}]}, Z_{N_i})\\
&= iH_i + jH_j.
\end{aligned}$$

In the above, the equality labeled $(a)$ follows from the conditional independence of $Z_{[N_i]}$ and $Z_{[N_i+1:N_{i+j}]}$ on $(X_{[i+1:i+j]}, Y_{[N_i+1:N_{i+j}]})$ and $(X_{[i]}, Y_{[N_i]}, Z_{[N_i-1]})$, respectively, given $N_i$. From Fekete's Lemma [22, Appendix II], this superadditivity proves the claim.

## APPENDIX V
## SPECIAL CASES OF $D_i^{\text{iud}}$

### A. *Bounds for $D_2^{\text{iud}}$*

It is easy to see that when $i = 2$, Equation (11) reduces to

$$\begin{aligned}
\mathfrak{H}_m^{(2)} &= H(Z_1 | Z_2 = -m, \mathcal{X}, \mathcal{Y})\\
&= \log_2(m+1) - \frac{1}{2^{m+1}} \sum_{x_{[m+1]}} h_2\Big(\frac{w(x_{[m+1]})}{m+1}\Big)\\
&= \log_2(m+1) - \frac{1}{2^{m+1}} \sum_{j=0}^{m+1} \binom{m+1}{j} h_2\Big(\frac{j}{m+1}\Big),
\end{aligned} \tag{15}$$

where $w(\cdot)$ denotes *Hamming weight*. Hence

$$D_2^{\text{iud}} = 1 - p - h_2(p) + (1-p)^3 \sum_{m\geq 0} (m+1)\mathfrak{H}_m^{(2)}. \tag{16}$$

For numerically estimating $\mathfrak{H}_m^{(2)}$ for large $m$, we can use the upper bound [56]

$$\binom{m+1}{j} \leq 2^{(m+1)h_2(\frac{j}{m+1})} \sqrt{\frac{m+1}{2\pi j(m+1-j)}}$$

to get a further lower bound[6] on $\mathfrak{H}_m^{(2)}$. On the other hand, to obtain a looser analytic lower bound, we can bound

$$\frac{1}{2^{m+1}} \sum_{j=0}^{m+1} \binom{m+1}{j} h_2\Big(\frac{j}{m+1}\Big) \leq 1 - 2^{-m},$$

to get $\mathfrak{H}_m^{(2)} \geq \log_2(m+1) - 1 + 2^{-m}$. This gives us

$$\begin{aligned}
D_2^{\text{iud}} \geq (1-p)^3 \Big(&\frac{4}{(2-p)^2} + \sum_{m\geq 0} (m+1)p^m \log_2(m+1)\Big)\\
&- h_2(p).
\end{aligned}$$

Unfortunately, it is not easy to evaluate the series

$$\vartheta = \sum_{m=2}^{\infty} m p^{m-1} \log_2 m$$

on the right hand side of the above inequality. Consider the function

$$f(x) = x p^{x-1} \ln x,$$

where $\ln(\cdot)$ is the natural logarithm. The $m^{\text{th}}$ term in the series can then be written as $(\log_2 e) f(m), m \geq 2$. It turns out that for

$$p < p^* \triangleq \exp\Big(-\frac{1+\ln 2}{2\ln 2}\Big) \approx 0.294832606,$$

we can lower bound the series $\vartheta$ by the integral

$$\begin{aligned}
\vartheta &\geq \log_2 e \int_2^{\infty} f(x)dx = \log_2 e \int_2^{\infty} x p^{x-1} \ln x dx\\
&= \frac{\log_2 e}{\ln p}\Big(\frac{p}{\ln p}(1+\ln 2) - 2p\ln 2 - \frac{1}{p}\mathrm{Ei}(2\ln p)\Big),
\end{aligned}$$

[6] We would like to get a lower bound on $D_2^{\text{iud}}$ since this will be a lower bound for $C^{\text{iud}}$ as well (cf. Equation (12)).

where $\mathrm{Ei}(x)$ is the *exponential integral* function defined as

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t} dt,$$

which can be numerically evaluated to arbitrary accuracy through a Taylor series expansion. Therefore, for $p < p^*$,

$$D_2^{\mathrm{iud}} \geq \frac{4(1-p)^3}{(2-p)^2} - h_2(p) + (1-p)^3 \frac{\log_2 e}{\ln p}\Big(\frac{p}{\ln p}(1+\ln 2) - 2p\ln 2 - \frac{1}{p}\mathrm{Ei}(2\ln p)\Big).$$

With $\mathfrak{H}_m^{(2)}$ as given in Equation (15), we can write

$$D_2^{\mathrm{iud}} = 1 + p\log_2 p - p\log_2(2e) + O(p^2)$$

for small $p$. This is loose compared to the bound obtained in [12]. This can be attributed to the fact that we evaluated $H(Z_1|Z_2, \mathcal{X}, \mathcal{Y})$ rather than $H(Z_1|\mathcal{X}, \mathcal{Y})$ to obtain $D_2^{\mathrm{iud}}$. In fact, this small-$p$ series expansion of $D_2^{\mathrm{iud}}$ is no better than that of the lower bound for the BDC in Corollary 12. We will improve this bound for small $p$ in the next subsection.

### B. Bounds for the case when $m = 1$

We now pursue the other case where (11) is easy to evaluate. Instead of evaluating $D_i^{\mathrm{iud}}$ exactly, we can further lower bound it as follows.

$$\begin{aligned}
D_i^{\mathrm{iud}} &= 1 - p - h_2(p) + (1-p)H(Z_1|Z_i, \mathcal{X}, \mathcal{Y}) \\
&= 1 - p - h_2(p) + (1-p) \\
&\quad \times \Big(\sum_{m\geq 0} \mathsf{P}(Z_i = -m) H(Z_1|Z_i = -m, \mathcal{X}, \mathcal{Y})\Big) \\
&= 1 - p - h_2(p) + (1-p) \\
&\quad \times \Big(\sum_{m=0}^{j} \mathsf{P}(Z_i = -m) H(Z_1|Z_i = -m, \mathcal{X}, \mathcal{Y})\Big) \\
&\triangleq 1 - p - h_2(p) + (1-p)\Psi_j^{(i)} \\
&\triangleq \mathfrak{D}_j^{(i)} \ \forall\ j \geq 0, i \geq 1.
\end{aligned}$$

We are essentially writing a series expansion for $D_i^{\mathrm{iud}}$ and lower bounding[7] it by the $j^{\mathrm{th}}$ partial sum. Note that we can write

$$\begin{aligned}
\Psi_j^{(i)} &= \Psi_{j-1}^{(i)} + \mathsf{P}(Z_i = -j)H(Z_1|Z_i = -j, \mathcal{X}, \mathcal{Y}) \\
&= \Psi_{j-1}^{(i)} + \psi_{i,j} p^j (1-p)^i
\end{aligned} \tag{17}$$

where $\psi_{i,m}$ was defined in Theorem 20. Clearly, the sequence $\{\Psi_j^{(i)}\}_{j\geq 0}$ is non-decreasing, and, in turn, so is the sequence $\{\mathfrak{D}_j^{(i)}\}_{j\geq 0}$. Since $\Psi_0^{(i)} = \psi_{i,0} = 0$, we have $\Psi_1^{(i)} = p(1-p)^i\psi_{i,1}$. Further, by definition,

$$D_i^{\mathrm{iud}} = \lim_{j\to\infty} \mathfrak{D}_j^{(i)} = \sup_{j\geq 0} \mathfrak{D}_j^{(i)}.$$

Thus for every $j \geq 0$, we can write

$$\begin{aligned}
C_{\mathrm{BDC}}^{\mathrm{iud}} &= \sup_{i\geq 1} D_i^{\mathrm{iud}} = \sup_{i\geq 1}\sup_{j\geq 0} \mathfrak{D}_j^{(i)} \overset{(a)}{=} \sup_{j\geq 0}\sup_{i\geq 1} \mathfrak{D}_j^{(i)} \\
&\geq \sup_{i\geq 1} \mathfrak{D}_j^{(i)} \triangleq \mathfrak{D}_j^{\mathrm{iud}},
\end{aligned}$$

where $(a)$ is true since $\mathfrak{D}_j^{(i)} \in [0,1]\ \forall\ i \geq 1, j \geq 0$.

From the channel model, $\mathsf{P}(X_{[i]} = x_{[i]}|Z_i = -1) = 2^{-i}$ since $\mathcal{X} \perp\!\!\!\perp \mathcal{Z}$ and $\mathcal{X}$ is i.u.d., and

$$\mathsf{P}(Y_{[i-1]} = y_{[i-1]}|X_{[i]} = x_{[i]}, Z_i = -1) = \frac{w_{y_{[i-1]}}(x_{[i]})}{i}.$$

For $y_{[i-1]} = x_{[i-1]-z_{[i-1]}}$ for some realization $z_{[i-1]}$ with the boundary conditions $z_0 = 0$ and $z_i = -1$,

$$\begin{aligned}
H(Z_1|Z_i = -1, &X_{[i]} = x_{[i]}, Y_{[i-1]} = y_{[i-1]}) \\
&= h_2\Big(\frac{1}{r_1(x_{[i]})}\Big)\mathbb{1}_{\mathcal{R}_1(x_{[i]}, y_{[i-1]})}
\end{aligned}$$

where $\mathcal{R}_1(x_{[i]}, y_{[i-1]})$ is the event that the single deletion occurred in the first run of $x_{[i]}$ to result in $y_{[i-1]}$. To see this, let $y_{[i-1]}$ represent a received word resulting from a single deletion upon transmission of $x_{[i]}$. Consider the two mutually exclusive and exhaustive cases in this scenario:

- The single deletion occurs in a run other than the first run of $x_{[i]}$. In this case, there is no ambiguity that $Z_1 = 0$, and the first run of $y_{[i-1]}$ is either the same or larger than[8] that of $x_{[i]}$.
- The single deletion occurs in the first run of $x_{[i]}$.
  - If $r_1(x_{[i]}) = 1$, there is no ambiguity that $Z_1 = -1$.
  - If $r_1(x_{[i]}) > 1$, the deleted symbol could be, with equal likelihood, one of the symbols comprising the first run of $x_{[i]}$. The uncertainty in $Z_1$ is $h_2\Big(\frac{1}{r_1(x_{[i]})}\Big)$.

  In both the above cases, the uncertainty can be written as $h_2\Big(\frac{1}{r_1(x_{[i]})}\Big)$.

Therefore,

$$\begin{aligned}
\psi_{i,1} &= i\sum_{x_{[i]}} \frac{1}{2^i} \sum_{y_{[i-1]}} \frac{w_{y_{[i-1]}}(x_{[i]})}{i} h_2\Big(\frac{1}{r_1(x_{[i]})}\Big)\mathbb{1}_{\mathcal{R}_1(x_{[i]}, y_{[i-1]})} \\
&= i\sum_{x_{[i]}} \frac{1}{2^i}\frac{r_1(x_{[i]})}{i} h_2\Big(\frac{1}{r_1(x_{[i]})}\Big) \\
&= \frac{1}{2^i}\sum_{j=1}^{i} j2^{i-j} h_2\Big(\frac{1}{j}\Big) + \frac{1}{2^i}h_2\Big(\frac{1}{i}\Big) \\
&= \sum_{j=1}^{i} \frac{j}{2^j} h_2\Big(\frac{1}{j}\Big) + \frac{1}{2^i}h_2\Big(\frac{1}{i}\Big) \\
&= \frac{1}{2}\sum_{j=1}^{i-2} \frac{j}{2^j}\log_2 j + 2\frac{i}{2^i}\log_2 i.
\end{aligned} \tag{18}$$

We observe that $\psi_{i,1}$ is non-decreasing in $i$, and converges exponentially to the value $\psi_1 \approx 1.288531275$. From (17) and (18), we have

$$\begin{aligned}
\mathfrak{D}_1^{(i)} &= 1 - p - h_2(p) + p(1-p)^{i+1}\psi_{i,1} \\
&= 1 + p\log_2 p - p\log_2(2e) + \psi_{i,1}p + O(p^2).
\end{aligned}$$

Since $D_i^{\mathrm{iud}} = \mathfrak{D}_1^{(i)} + \sum_{j\geq 2} p^j(1-p)^{i+1}\psi_{i,j}$, we have

$$D_i^{\mathrm{iud}} = 1 + p\log_2 p - p\log_2(2e) + \psi_{i,1}p + O(p^2).$$

[7] All terms in the series expansion are non-negative.

[8] When the second run of $x_{[i]}$ disappears.

Thus, from Equation (12)

$$C_{\rm BDC}^{\rm iud} = 1 + p\log_2 p - \mathtt{d}p + O(p^2) \tag{19}$$

where $\mathtt{d} = \log_2(2e) - \psi_1 \approx 1.154163765$. We note here that this is exactly the same bound obtained in [12] with a completely different technique. Since this bound was shown to be tight for small $p$, we have that the capacity of the BDC itself is given by the above expression for small $p$.

*Discussion :* The advantage in the evaluation of the above bound was that, when we restrict to the case of a single deletion, the ambiguity in the first channel state $Z_1$ arises only when $r_1(x_{[i]}) > 1$, in which case the uncertainty is exactly $h_2\Big(\frac{1}{r_1(x_{[i]})}\Big)$. This, however, is not true when there are 2 or more deletions, wherein we will have to count subsequence weights of sequences.

### C. Bounds for Markov-1 rates

We can get similar bounds as in the previous two subsections for first-order Markov inputs. Further, since the channel has no bias for the input symbols, we can confine our attention to *symmetric* Markov inputs. But these calculations will have to keep track of *ascents* and *descents* in sequences, and are therefore more tedious.

Proceeding along the same lines as in Appendix V-A, we can write for $\mathsf{P}(X_i = x\oplus 1 | X_{i-1} = x) = \alpha \in [0,1]$,

$$D_2^{\mathcal{M}1} = \Big[\max_\alpha \Big(h_2(\alpha) + (1-p)^2\sum_{m\geq 0}(m+1)p^m \ell_m(\alpha)\Big)\Big] \times (1-p) - h_2(p), \text{ where}$$

$$\ell_m(\alpha) = \log_2(m+1) - \sum_{j=0}^{m+1} h_2\Big(\frac{j}{m+1}\Big)\eta(\alpha, j, m+1),$$

and $\eta(\cdot)$ is defined recursively as

$$\begin{aligned}\eta(\alpha,j,m) &= \eta_0(\alpha,j,m) + \eta_1(\alpha,j,m)\\ \eta_0(\alpha,j,m) &= (1-\alpha)\eta_0(\alpha,j,m-1)\\ &\quad + \alpha\eta_1(\alpha,j,m-1)\\ \eta_1(\alpha,j,m) &= (1-\alpha)\eta_1(\alpha,j-1,m-1)\\ &\quad + \alpha\eta_0(\alpha,j-1,m-1)\end{aligned}$$

with $\eta_k(\alpha, km, m) = \frac{1}{2}(1-\alpha)^{m-1}$, $\eta_k(\alpha,(1-k)m,m) = 0$ and $\eta_k(\alpha,j,m) = 0\ \forall\ j\notin[m]$ for $k\in\{0,1\}$.

Similarly, we can also evaluate

$$\begin{aligned}\mathfrak{D}_1^{\mathcal{M}1} &= -h_2(p) + (1-p)\times\\ &\max_\alpha\Big[h_2(\alpha) + p\cdot\sup_{i\geq 1}(1-p)^i\Big(\alpha\sum_{j=1}^{i} j(1-\alpha)^{j-1}h_2\Big(\frac{1}{j}\Big)\\ &\qquad + i(1-\alpha)^i h_2\Big(\frac{1}{i}\Big)\Big)\Big].\end{aligned}$$

However, both $D_2^{\mathcal{M}1}$ and $\mathfrak{D}_1^{\mathcal{M}1}$ turn out to be better than their SIR counterparts by less than 2%.

*Discussion :* Although first-order Markov inputs are expected to perform better than i.u.d. inputs, we see that the bounds we obtained are almost the same in the two cases. This is because we are considering two special cases, the first when $i = 2$ wherein all but a single symbol were deleted, and the second when $m = 1$ wherein a single symbol was deleted; and in these cases, a first-order Markov input is not significantly different than i.u.d. inputs. □

## Appendix VI
## Proofs of Results for the BRC

### A. Proof of Corollary 24

We have from Proposition 16 and Lemma 17 that

$$\begin{aligned}C_{\rm BRC}^{\mathcal{M}1} \geq R_2^{\mathcal{M}1} \triangleq \max_\alpha\Big[h_2(\alpha) + \frac{H(Z_1|Z_2,\mathcal{X},\mathcal{Y})}{1-p}\\ - \frac{p+(1-\alpha)(1-p)}{1-p}h_2\Big(\frac{p}{p+(1-\alpha)(1-p)}\Big)\Big]\end{aligned}$$

where we have used the expression for $H(Z_1|\mathcal{Y})$ from the proof of Theorem 23. Observe that $Z_2 \in\{0,1,2\}$, and among these possibilities, the only event wherein there is an ambiguity in the value of $Z_1$ is when $Z_2 = 1$. Thus, we can see easily that $H(Z_1|Z_2,\mathcal{X},\mathcal{Y}) = 2p(1-p)(1-\alpha)$. Hence

$$\begin{aligned}R_2^{\mathcal{M}1} = \max_\alpha\Big[h_2(\alpha) + 2p(1-\alpha)\\ - \frac{p+(1-\alpha)(1-p)}{1-p}h_2\Big(\frac{p}{p+(1-\alpha)(1-p)}\Big)\Big].\end{aligned}$$

It can be shown that the optimal $\alpha$ in the above is given by

$$\alpha^* = \frac{1}{(1-p)(2^{2p}+1)}.$$

Note that $\alpha^*$ is always larger than $\frac{1}{2}$, and $\alpha^* \leq 1$ for $p\leq p_*$ where

$$p_* \approx 0.734675821.$$

Plugging this back in the expression for $R_2^{\mathcal{M}1}$ ends the proof.

### B. Proof of Corollary 25

From Proposition 16 and Lemma 17, for i.u.d. inputs,

$$\begin{aligned}C_{\rm BRC}^{\rm iud} &= 1 - \frac{H(Z_1|\mathcal{Y})}{1-p} + \sup_{i\geq 1}\frac{H(Z_1|Z_i,\mathcal{X},\mathcal{Y})}{1-p}\\ &= 1 - \frac{1+p}{2(1-p)}h_2\Big(\frac{2p}{1+p}\Big)\\ &\quad + \sup_{i\geq 1}\frac{\Big(\sum_{m=0}^{i}\mathsf{P}(Z_i = m)H(Z_1|Z_i = m,\mathcal{X},\mathcal{Y})\Big)}{1-p}\\ &= 1 - \frac{1+p}{2(1-p)}h_2\Big(\frac{2p}{1+p}\Big)\\ &\quad + \sup_{i\geq 1}\Big(\sum_{m=0}^{i}\binom{i}{m}p^m(1-p)^{i-m-1}\\ &\qquad\qquad \times H(Z_1|Z_i = m,\mathcal{X},\mathcal{Y})\Big)\\ &= 1 - \frac{1+p}{2(1-p)}h_2\Big(\frac{2p}{1+p}\Big)\\ &\quad + \sup_{i\geq 1}\Big(ip(1-p)^{i-2}H(Z_1|Z_i = 1,\mathcal{X},\mathcal{Y})\Big) + O(p^2),\end{aligned}$$

where we have used the expression for $H(Z_1|\mathcal{Y})$ from the proof of Theorem 23 for $\alpha = \frac{1}{2}$. The last equality is true since $H(Z_1|Z_i = 0, \mathcal{X}, \mathcal{Y}) = 0$.

As shown in the proof of Theorem 23, we can write

$$\begin{aligned}H(Z_1|Z_i = 1,&\mathcal{X}, \mathcal{Y})\\ &= \mathsf{E}[h_2(\mathsf{P}(Z_1 = 0|Z_i = 1, r_1^0(x_{\mathbb{N}}), r_1^0(y_{\mathbb{N}})))].\end{aligned}$$

Further, there is no ambiguity in $Z_1$ if the single replication does not occur in the first run of $x_{\mathbb{N}}$. Therefore, for a first-order Markov input process,

$$\begin{aligned}H(Z_1|Z_i = 1, \mathcal{X}, \mathcal{Y}) &= \mathsf{E}[h_2(\mathsf{P}(Z_1 = 0|Z_i = 1,\\ &\qquad r_1^0(x_{\mathbb{N}}) = l, r_1^0(y_{\mathbb{N}}) = l+1))]\\ &= \sum_{l=1}^{i-1}(1-\alpha)^l\alpha\frac{l+1}{i}h_2\Big(\frac{1}{l+1}\Big) + (1-\alpha)^i h_2\Big(\frac{1}{i}\Big).\end{aligned}$$

For $\alpha = \frac{1}{2}$, we get

$$iH(Z_1|Z_i = 1, \mathcal{X}, \mathcal{Y}) = \frac{1}{2}\sum_{l=1}^{i-2}\frac{l}{2^l}\log_2 l + 2\frac{i}{2^i}\log_2 i = \psi_{i,1}$$

from Equation(18). Thus,

$$\begin{aligned}C_{\mathrm{BRC}}^{\mathrm{iud}} &= 1 - \frac{1+p}{2(1-p)}h_2\Big(\frac{2p}{1+p}\Big)\\ &\quad + \sup_{i\geq 1}\Big(p(1-p)^{i-2}\psi_{i,1}\Big) + O(p^2),\\ &= 1 + p\log_2 p + \log_2\Big(\frac{2}{e}\Big)p\\ &\quad + \sup_{i\geq 1}\Big(p(1-p)^{i-2}\psi_{i,1}\Big) + O(p^2),\\ &= 1 + p\log_2 p + \mathtt{r}p + O(p^2),\end{aligned}$$

where $\mathtt{r} = \log_2(\frac{2}{e}) + \psi_1 = 2 - \mathtt{d} \approx 0.845836235$. As was the case for the BDC, we expect this to be a tight bound for the capacity for small $p$.

## Appendix VII
## Proof of Lemma 26

From the definition of the $\mathcal{Z}^{(m)}$ process in Equation (14), it is clear that $Z_i^{(m)} \in \{-m, -m+1, \cdots, -1, 0, 1, 2, \cdots, m\}$ for every $i \in \mathbb{N}$, and that it is a Markov chain. It is therefore a finite Markov chain. The time-inhomogeneity follws by noting the transition probabilities between states, which can be easily shown to be given as follows. For $-m < j < m$, $i \geq 1$,

$$\begin{aligned}\mathsf{P}(Z_i^{(m)} &= k|Z_{i-1}^{(m)} = j)\\ &= \begin{cases} p_{\mathsf{r}}, & k = j+1\\ (1-p_{\mathsf{r}})(1-p_{\mathsf{d}})p_{\mathsf{d}}^{j-k}, & -m < k \leq j\\ (1-p_{\mathsf{r}})p_{\mathsf{d}}^{j+m}, & k = -m\\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

From the states $\{\pm m\}$, the transition probabilities are

$$\begin{aligned}\mathsf{P}(Z_i^{(m)} &= k|Z_{i-1}^{(m)} = -m)\\ &= \begin{cases} 1 - p_{\mathsf{r}}\underline{\mathsf{p}}(i,m), & k = -m\\ p_{\mathsf{r}}\underline{\mathsf{p}}(i,m), & k = -m+1\\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

where

$$\underline{\mathsf{p}}(i,m) = \frac{\mathsf{P}(Z_{i-1} = -m)}{\mathsf{P}(Z_{i-1} \leq -m)},$$

and for $i$ such that $\mathsf{P}(Z_{i-1} \geq m) > 0$,

$$\begin{aligned}\mathsf{P}(Z_i^{(m)} &= k|Z_{i-1}^{(m)} = m)\\ &= \begin{cases} 1 - p_{\mathsf{d}}(1-p_{\mathsf{r}})\overline{\mathsf{p}}(i,m,p_{\mathsf{d}}), & k = m\\ (1-p_{\mathsf{d}})(1-p_{\mathsf{r}})p_{\mathsf{d}}^{m-k}\overline{\mathsf{p}}(i,m,p_{\mathsf{d}}), & -m < k < m\\ (1-p_{\mathsf{r}})p_{\mathsf{d}}^{2m}\overline{\mathsf{p}}(i,m,p_{\mathsf{d}}), & k = -m\\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

where

$$\overline{\mathsf{p}}(i,m,p) = \frac{\sum_{l=0}^{\infty}\mathsf{P}(Z_{i-1} = m+l)p^l}{\mathsf{P}(Z_{i-1} \geq m)}.$$

Note that it is only transitions from the boundary states $\{\pm m\}$ that have time-dependent probabilities.

As was noted in the proofs of Lemmas 6 and 15, we can write the $\mathcal{Z}$ process as $Z_n = \sum_{i=1}^n \Xi_i$ where $\{\Xi_i\}_{i\geq 1}$ is an i.i.d. process with

$$\mathsf{P}(\Xi_1 = \xi) = \begin{cases} p_{\mathsf{r}}, & \xi = 1\\ (1-p_{\mathsf{d}})(1-p_{\mathsf{r}})p_{\mathsf{d}}^{-\xi}, & \forall\ \xi \leq 0\\ 0, & \text{otherwise.}\end{cases}$$

As noted before, from the SLLN, $\frac{Z_n}{n} \to \mathsf{E}[\Xi_1] = \frac{p_{\mathsf{r}} - p_{\mathsf{d}}}{1-p_{\mathsf{d}}} \triangleq \chi$ a.s. as $n \to \infty$. Let us write

$$\mathsf{Var}[\Xi_1] = \mathsf{E}[\Xi_1^2] - \mathsf{E}[\Xi_1]^2 = \frac{p_{\mathsf{r}} + p_{\mathsf{d}} + p_{\mathsf{d}}^2 - 3p_{\mathsf{d}}p_{\mathsf{r}}}{(1-p_{\mathsf{d}})^2} \triangleq \nu^2.$$

From the central limit theorem (CLT), we have

$$\begin{aligned}\mathsf{P}(Z_n \geq m) &= \mathsf{P}\Big(\frac{Z_n - n\chi}{\sqrt{n}\nu} \geq \frac{m - n\chi}{\sqrt{n}\nu}\Big) \stackrel{n\to\infty}{\longrightarrow} \mathsf{Q}\Big(\frac{\mathtt{t}}{\nu}\Big),\\ \mathsf{P}(Z_n \leq -m) &= \mathsf{P}\Big(\frac{Z_n - n\chi}{\sqrt{n}\nu} \leq \frac{-m - n\chi}{\sqrt{n}\nu}\Big) \stackrel{n\to\infty}{\longrightarrow} \mathsf{Q}\Big(\frac{\mathtt{b}}{\nu}\Big),\end{aligned}$$

where

$$\mathtt{t} \triangleq \lim_{n\to\infty}\frac{m - n\chi}{\sqrt{n}} \quad \text{and} \quad \mathtt{b} \triangleq \lim_{n\to\infty}\frac{m + n\chi}{\sqrt{n}}$$

with $\mathtt{t}, \mathtt{b} \in \mathbb{R} \cup \{\pm\infty\}$, and

$$\mathsf{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-\frac{u^2}{2}}du.$$

Writing $\mathtt{m} \triangleq \lim_{n\to\infty}\frac{m}{n}$, we can say that when $\chi > 0$,

$$\mathsf{P}(Z_n^{(m)} = m) = \mathsf{P}(Z_n \geq m) \stackrel{n\to\infty}{\longrightarrow} \begin{cases} 1, & \mathtt{m} < \chi\\ \frac{1}{2}, & \mathtt{m} = \chi\\ 0, & \mathtt{m} > \chi,\end{cases}$$

and when $\chi < 0$,

$$\mathsf{P}(Z_n^{(m)} = -m) = \mathsf{P}(Z_n \leq -m) \stackrel{n\to\infty}{\longrightarrow} \begin{cases} 1, & \mathtt{m} < -\chi\\ \frac{1}{2}, & \mathtt{m} = -\chi\\ 0, & \mathtt{m} > -\chi.\end{cases}$$

When $\chi = 0$, for $m = o(\sqrt{n})$,

$$\left.\begin{aligned}\mathsf{P}(Z_n^{(m)} = m)\\ \mathsf{P}(Z_n^{(m)} = -m)\end{aligned}\right\} \stackrel{n\to\infty}{\longrightarrow} \frac{1}{2}.$$

Hence

$$\mathsf{P}(Z_n^{(m)} \in \{\pm m\}) \stackrel{n\to\infty}{\longrightarrow} 1 \begin{cases} \text{for } m = o(n) \text{ when } \chi \neq 0, \\ \text{for } m = o(\sqrt{n}) \text{ when } \chi = 0. \end{cases}$$

The result then follows by noting that $\chi$ is equal (or not equal) to 0 accordingly as $p_{\mathsf{d}}$ is equal (or not equal, respectively) to $p_{\mathsf{r}}$.

## Appendix VIII
## Proof of Lemma 27

We first note that although $N_n$ and $N_n^{(m)}$ might themselves differ, both sets $\{\Gamma_{[N_n]}\}$ and $\{\Gamma^{(m)}_{[N_n^{(m)}]}\}$ are subsets of $[n]\cup\{0\}$. Therefore, assuming that all random variables $X_i$ where $i \notin [n]$ are constants (in particular, we assume that these random variables are all equal to 0), we can consider the above sets of indices to be $\{\Gamma_{[N(m,n)]}\}$ and $\{\Gamma^{(m)}_{[N(m,n)]}\}$ respectively, where we define

$$N(m,n) \triangleq N_n \vee N_n^{(m)} = \max\{N_n, N_n^{(m)}\}.$$

We have $N(m,n) < \infty$ a.s. for every $n \in \mathbb{N}, m \in \mathbb{Z}^+$.

Let $S_1^+ = \inf\{i > 0 : Z_{i+1} > m\}$ and $T_1^+ = \inf\{i > 0 : Z_{S^+ +1+i} \leq m\}$, where we define $\inf\emptyset = \infty$. Then, $\{S_1^+ + 1, S_1^+ + 2, \cdots, T_1^+\} \cap [N(m,n)]$ is the set of instances where $Z_i$ and $Z_i^{(m)}$ differ for the first time as a result of $Z_i$ exceeding $m$. In this case, $Z_{S_1^+} = m, Z_{S_1^+ +1} = m+1$ with probability 1 from the definition of the $\mathcal{Z}$ process. Further, $m+1 \leq Z_{S_1^+ + j} \leq m+j, j = 1, 2, \cdots, T_1^+$ a.s., implying for this range of $j$s that $S_1^+ - m \leq \Gamma_{S_1^+ + j} \leq S_1^+ + j - m - 1$ a.s.. But, by definition, $Z^{(m)}_{S_1^+ + j} = m, j = 0, 1, \cdots, T_1^+$, and hence $\Gamma^{(m)}_{S_1^+ + j} = S_1^+ + j - m$. Thus, if we write $\mathbb{U}_1^+ = \{S_1^+, S_1^+ + 1, \cdots, S_1^+ + T_1^+\}$, then we have

$$\{\Gamma_{\mathbb{U}_1^+}\} \subset \{\Gamma^{(m)}_{\mathbb{U}_1^+}\} \text{ a.s..}$$

Since $Z_{S_1^+} = Z^{(m)}_{S_1^+} = m, \Gamma_i \leq \Gamma_{S_1^+} = S_1^+ - m\ \forall\ i \leq S_1^+$ a.s. from Lemma 7. Similarly, since $Z_{S_1^+ + T_1^+ + 1} = Z^{(m)}_{S_1^+ + T_1^+ + 1} \leq m, \Gamma_i \geq \Gamma_{S_1^+ + T_1^+ + 1} \geq S_1^+ + T_1^+ + 1 - m\ \forall\ i \geq S_1^+ + T_1^+ + 1$ a.s. from Lemma 7. It follows that the indices in $\{\Gamma^{(m)}_{\mathbb{U}_1^+}\} \setminus \{\Gamma_{\mathbb{U}_1^+}\}$ cannot appear in $\{\Gamma_{\mathbb{U}}\}$ for any $\mathbb{U} \subset [N(m,n)] \setminus \mathbb{U}_1^+$. Using similar arguments, by recursively defining for $i \geq 2$

$$\begin{aligned} S_i^+ &= \inf\{i > S_{i-1}^+ + T_{i-1}^+ : Z_{i+1} > m\}, \\ T_i^+ &= \inf\{i > 0 : Z_{S_i^+ + 1 + i} \leq m\}, \end{aligned}$$

and letting $\mathbb{U}_i^+ = \{S_i^+, S_i^+ + 1, \cdots, S_i^+ + T_i^+\}$, we can show that

$$\{\Gamma_{\mathbb{U}_i^+}\} \subset \{\Gamma^{(m)}_{\mathbb{U}_i^+}\} \text{ a.s.} \forall\ i \geq 1.$$

Similarly, consider

$$\begin{aligned} S_0^- &= T_0^- = 0, \\ S_i^- &= \inf\{i > S_{i-1}^- + T_{i-1}^- : Z_i < -m\} \text{ and} \\ T_i^- &= \inf\{i > 0 : Z_{S_i^- + i} \geq -m\}, i \geq 1. \end{aligned}$$

Then, $Z_{S_i^- + T_i^-} = -m$ and $Z_{S_i^- + T_i^- - 1} = -m-1$ with probability 1. Further, with probability 1, $-m - j \leq Z_{S_i^- + T_i^- - j} \leq -m - 1, j = 1, 2, \cdots, T_i^-$ a.s., implying $S_i^- + T_i^- - j + m + 1 \leq \Gamma_{S_i^- + T_i^- - j} \leq S_i^- + T_i^- + m$ a.s.. By definition, $\Gamma^{(m)}_{S_i^- + T_i^- - j} = S_i^- + T_i^- - j + m, j = 0, 1, \cdots, T_i^-$, and consequently

$$\{\Gamma_{\mathbb{U}_i^-}\} \subset \{\Gamma^{(m)}_{\mathbb{U}_i^-}\} \text{ a.s.} \forall\ i \geq 1$$

where $\mathbb{U}_i^- = \{S_i^-, S_i^- + 1, \cdots, S_i^- + T_i^-\}$. As before, the missing indices cannot appear in $\Gamma_{\mathbb{U}}$ for any $\mathbb{U} \subset [N(m,n)] \setminus \mathbb{U}_i^-$.

Therefore, writing $\mathbb{U}^\pm = \bigcup_{i\geq 1}(\mathbb{U}_i^+ \cup \mathbb{U}_i^-)$, we have

$$\{\Gamma_{\mathbb{U}^\pm}\} \subset \{\Gamma^{(m)}_{\mathbb{U}^\pm}\} \text{ a.s.}$$

Let $\mathbb{U}^0 \triangleq [N(m,n)] \setminus \mathbb{U}^\pm$. Since $\mathbb{U}^\pm$ consists of all indices $i$ where $\Gamma_i$ and $\Gamma_i^{(m)}$ differ, we have from the above relation that almost surely,

$$\begin{aligned} \{\Gamma_{\mathbb{U}^\pm}\} \cup \{\Gamma_{\mathbb{U}^0}\} &\subset \{\Gamma^{(m)}_{\mathbb{U}^\pm}\} \cup \{\Gamma^{(m)}_{\mathbb{U}^0}\} \\ \Rightarrow \{\Gamma_{[N(m,n)]}\} &\subset \{\Gamma^{(m)}_{[N(m.n)]}\} \\ \Rightarrow \{\Gamma_{[N_n]}\} \cap [n] &\subset \{\Gamma^{(m)}_{[N_n^{(m)}]}\} \cap [n]. \end{aligned}$$

We are interested in the intersection in the last step above since only indices in the set $[n]$ are indices of non-constant random variables.

We can use an argument similar to the one above to show that $\forall\ m \in \mathbb{Z}^+$,

$$\{\Gamma^{(m+1)}_{[N_n^{(m+1)}]}\} \cap [n] \subset \{\Gamma^{(m)}_{[N_n^{(m)}]}\} \cap [n] \text{ a.s..}$$

## Appendix IX
## Proof of Proposition 28

We use the result from Lemma 27. Let us define

$$\begin{aligned} \mathbb{S}_m &= \bigcap_{n\geq 1} \{\varsigma \in \mathbb{S} : \{\Gamma_{[N_n]}\} \cap [n] \subset \{\Gamma^{(m)}_{[N_n^{(m)}]}\} \cap [n]\}, \text{ and} \\ \hat{\mathbb{S}}_m &= \bigcap_{n\geq 1} \{\varsigma \in \mathbb{S} : \{\Gamma^{(m+1)}_{[N_n^{(m+1)}]}\} \cap [n] \subset \{\Gamma^{(m)}_{[N_n^{(m)}]}\} \cap [n]\} \end{aligned}$$

and let

$$\mathbb{S}^* = \bigcap_{m\in\mathbb{Z}^+} (\mathbb{S}_m \cap \hat{\mathbb{S}}_m).$$

Clearly, $\mathsf{P}(\mathbb{S}^*) = 1$. Then, confining the expectations over the set $\mathbb{S}^*$,

$$\begin{aligned} I(X_{[n]}; Y^{(m)}_{[N_n^{(m)}]}) &- I(X_{[n]}; Y_{[N_n]}) \\ &= I_{\mathbb{S}^*}(X_{[n]}; X_{\Gamma^{(m)}_{[N_n^{(m)}]} \setminus \Gamma_{[N_n]}} | X_{\Gamma_{[N_n]}}) \geq 0, \end{aligned}$$

where $I_{\mathbb{S}^*}(\cdot)$ denotes the mutual information obtained after confining the expectations to the set $\mathbb{S}^*$. Similarly, we have $I(X_{[n]}; Y^{(m+1)}_{[N_n^{(m+1)}]}) \leq I(X_{[n]}; Y^{(m)}_{[N_n^{(m)}]})$.

## Appendix X
## Extending the Measures $\mathsf{P}_{\langle m\rangle}$ to $\mathscr{B}$

We will first assume that $\mathbb{S} = \mathbb{S}_{\mathcal{X}} \times \mathbb{S}_{\mathcal{Z}}$ and that $\mathscr{B} = \mathscr{B}_{\mathcal{X}} \times \mathscr{B}_{\mathcal{Z}}$ with $\mathscr{B}_{\mathcal{X}} = \sigma(\mathcal{X})$ and $\mathscr{B}_{\mathcal{Z}} = \sigma(\mathcal{Z})$, i.e., the space $(\mathbb{S}, \mathscr{B})$ is a product space. Since in our model $\mathcal{X} \perp \mathcal{Z}$, there is no loss of generality in this assumption.

By defining the stationary transition probabilities $\mathsf{P}_{\langle m\rangle}(Z_1|Z_0)$ as in Section V-B, the measures $\mathsf{P}_{\langle m\rangle}$ are well-defined over $\mathscr{G}_m = \sigma(\mathcal{Z}^{(m)})$. Let $(\mathcal{Z}^{(m)})^{-1}(\overline{z}) \triangleq \{\varsigma \in \mathbb{S}_{\mathcal{Z}} : \mathcal{Z}^{(m)}(\varsigma) = \overline{z}\}$ for $\overline{z} \in \overline{\mathbb{Z}}_{\pm m}$, and similarly $\mathcal{Z}^{-1}(\overline{z}) \triangleq \{\varsigma \in \mathbb{S}_{\mathcal{Z}} : \mathcal{Z}(\varsigma) = \overline{z}\}$. Then, clearly

$$\mathcal{Z}^{-1}(\overline{z}) \subset (\mathcal{Z}^{(m)})^{-1}(\overline{z}) \;\forall\; \overline{z} \in \overline{\mathbb{Z}}_{\pm m}$$

and

$$\mathcal{Z}^{-1}(\overline{z}) \in \mathscr{B}_{\mathcal{Z}}, (\mathcal{Z}^{(m)})^{-1}(\overline{z}) \in \mathscr{G}_m \;\forall\; \overline{z} \in \overline{\mathbb{Z}}_{\pm m}.$$

Then, we define

$$\mathsf{P}_{\langle m\rangle}(\mathcal{Z}^{-1}(\overline{z})) = \mathsf{P}_{\langle m\rangle}((\mathcal{Z}^{(m)})^{-1}(\overline{z})) \;\forall\; \overline{z} \in \overline{\mathbb{Z}}_{\pm m}. \tag{20}$$

This will imply that for every $\overline{z} \in \overline{\mathbb{Z}}_{\pm m}$,

$$\mathsf{P}_{\langle m\rangle}((\mathcal{Z}^{(m)})^{-1}(\overline{z}) \setminus \mathcal{Z}^{-1}(\overline{z})) = 0.$$

By definition, we also have for $\overline{z} \in \overline{\mathbb{Z}} \setminus \overline{\mathbb{Z}}_{\pm m}$ that $(\mathcal{Z}^{(m)})^{-1}(\overline{z}) = \emptyset$ so that the associated probability is zero under any measure $\mathsf{P}, \mathsf{P}_{\langle m\rangle}$. We can now consider the space $(\mathbb{S}_{\mathcal{Z}}, \mathscr{B}_{\mathcal{Z}}, \mathsf{P}_{\langle m\rangle})$ to be obtained from $(\mathbb{S}_{\mathcal{Z}}, \mathscr{G}_m, \mathsf{P}_{\langle m\rangle})$ along with the definition (20) and subsequent *completion* [57, §2.6.19].

By now defining $\mathsf{P}_{\langle m\rangle}(\mathcal{X}) = \mathsf{P}(\mathcal{X})$ independent of $m$, we can extend the measure $\mathsf{P}_{\langle m\rangle}$ to $\mathscr{B} = \sigma(\{\mathcal{X}, \mathcal{Z}\})$ for each $m \in \mathbb{Z}^+$ as required.

## Appendix XI
## Proof of Lemma 37

As noted in the proof of Lemma 26, we have for every $n \in \mathbb{N}$ that $Z_n$ is the $n^{\text{th}}$ partial sum of the i.i.d. process $\{\Xi_i\}_{i\geq 1}$. For the SDRC, we have $\mathsf{E}[\Xi_1] = \chi = 0$ and $\mathsf{Var}[\Xi_1] = \nu^2 = \frac{2p}{1-p} < \infty$ since $p \in [0, 1)$.

Let $\mathscr{S}_n = \sigma(\{Z_n\}) \subset \mathscr{B}$, the sigma-algebra generated by $Z_n$, for every $n \in \mathbb{N}$. Clearly, $\mathscr{S}_n = \sigma(\{\Xi_{[n]}\})$ so that $\{\mathscr{S}_n\}_{n\geq 1}$ is a filtration, and $Z_n \in \mathscr{S}_n$ by definition. Let $\mathscr{S}_n \uparrow \mathscr{S} \subset \mathscr{B}$ as $n \to \infty$. Then for every $n \in \mathbb{N}$, $Z_n \in L^2(\mathbb{S}, \mathscr{B}, \mathsf{P})$ since

$$\begin{aligned}\mathsf{E}[|Z_n|^2] = \mathsf{E}[Z_n^2] &= \mathsf{E}\Big[\Big(\sum_{i=1}^{n} \Xi_i\Big)^2\Big] \\ &= \sum_{i=1}^{n} \mathsf{E}[\Xi_i^2] + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathsf{E}[\Xi_i \Xi_j] \\ &= n \cdot \mathsf{Var}[\Xi_1] + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathsf{E}[\Xi_i]\mathsf{E}[\Xi_j] \\ &= n\frac{2p}{1-p} < \infty.\end{aligned}$$

Further, $\mathsf{E}[Z_n|\mathscr{S}_{n-1}] = \mathsf{E}[Z_{n-1} + \Xi_n|\mathscr{S}_{n-1}] = Z_{n-1}$. Therefore, $\{Z_n, \mathscr{S}_n\}_{n\geq 1}$ is a martingale under the measure $\mathsf{P}$. Consequently, $\{|Z_n|, \mathscr{S}_n\}_{n\geq 1}$ is a submartingale.

Since $|Z_n| \in L^2(\mathbb{S}, \mathscr{B}, \mathsf{P})$, from Doob's submartingale inequality [20, §14.6], we have

$$\mathsf{P}(\max_{i=1}^{n} |Z_i| \geq m) \leq \frac{\mathsf{E}[|Z_n|^2]}{m^2} = \Big(\frac{2p}{1-p}\Big)\frac{n}{m^2}.$$

We have (cf. Section V-A) that $N_n^{(m)} \leq n + m$. The result then follows by noting that

$$\mathsf{P}\Big(\max_{i=1}^{N_n^{(m)}} |Z_i| \geq m\Big) \leq \mathsf{P}(\max_{i=1}^{n+m} |Z_i| \geq m)$$

and the above result. The bound with respect to the measure $\mathsf{P}_{\langle m\rangle}$ is true because

$$\begin{aligned}&\mathsf{P}_{\langle m\rangle}\Big(\vartheta \in \mathbb{S} : \max_{i=1}^{N_n^{(m)}(\vartheta)} |Z_i(\vartheta)| \geq m\Big) \\ &\qquad = \mathsf{P}\Big(\vartheta \in \mathbb{S} : \max_{i=1}^{N_n^{(m)}(\vartheta)} |Z_i(\vartheta)| \geq m\Big)\end{aligned}$$

from the definition of the measure $\mathsf{P}_{\langle m\rangle}$ (See Section V-B and Appendix X).

## Appendix XII
## Proof of Proposition 38

We start with a small Lemma.

*Lemma 39:* Let $(\mathbb{T}, \mathscr{A})$ be a measurable space, and let $\{\mathsf{Q}_n\}_{n\geq 1}$, $\mathsf{Q}$ all be probability measures on this space. Suppose that

i) For every $n \geq 1$, there is a set $\mathbb{B}_n \in \mathscr{A}$ such that $\mathsf{Q}_n(\mathbb{A}) = \mathsf{Q}(\mathbb{A})$ for every $\mathbb{A} \subset \mathbb{B}_n$, $\mathbb{A} \in \mathscr{A}$.
ii) $\mathsf{Q}(\mathbb{B}_n) \to 1$ as $n \to \infty$.

Then the measures $\mathsf{Q}_n$ converge in total variation to $\mathsf{Q}$, i.e., $\mathsf{Q}_n \xrightarrow{tv} \mathsf{Q}$ as $n \to \infty$.

*Proof:* From ii), for every $\epsilon > 0$, there exists $n'(\epsilon) \in \mathbb{N}$ such that

$$\mathsf{Q}(\mathbb{B}_n) \geq 1 - \epsilon \;\forall\; n \geq n'(\epsilon).$$

From i), $\mathsf{Q}_n(\mathbb{A} \cap \mathbb{B}_n) = \mathsf{Q}(\mathbb{A} \cap \mathbb{B}_n)$ for every $n \geq 1$, $\mathbb{A} \in \mathscr{A}$. Therefore, for every $\epsilon > 0$,

$$\begin{aligned}||\mathsf{Q}_n - \mathsf{Q}|| &= 2 \sup_{\mathbb{A}\in\mathscr{A}} |\mathsf{Q}_n(\mathbb{A}) - \mathsf{Q}(\mathbb{A})| \\ &= 2 \sup_{\mathbb{A}\in\mathscr{A}} |\mathsf{Q}_n(\mathbb{A} \cap \mathbb{B}_n^{\mathsf{C}}) - \mathsf{Q}(\mathbb{A} \cap \mathbb{B}_n^{\mathsf{C}})| \\ &\leq 2\epsilon \;\forall\; n \geq n'(\epsilon).\end{aligned}$$

Hence $\mathsf{Q}_n \xrightarrow{tv} \mathsf{Q}$ as $n \to \infty$. ∎

Note that

$$\mathbb{D}_{n,m} \triangleq \Big\{\vartheta \in \mathbb{S} : \max_{i=1}^{N_n^{(m)}(\vartheta)} |Z_i(\vartheta)| \geq m\Big\}$$

is the subset of $\mathbb{S}$ in $\mathscr{B}$ where $\mathsf{P}_{\langle m\rangle}(X_{[n]}, Y^{(m)}_{[N_n^{(m)}]})$ differs from $\mathsf{P}(X_{[n]}, Y_{[N_n]})$. From Lemma 37, we have

$$\mathsf{P}_{\langle m(n)\rangle}(\mathbb{D}_{n,m(n)}) = \mathsf{P}(\mathbb{D}_{n,m(n)}) \to 0$$

as $n \to \infty$, whenever $m(n) = \omega(\sqrt{n})$. Consider henceforth that $m(n)$ satisfies this condition. In Lemma 39 above, set $\mathbb{T} = \mathbb{S}$, $\mathscr{A} = \mathscr{B}$, $\mathsf{Q}_n = \mathsf{P}_{\langle m(n)\rangle}$ and $\mathsf{Q} = \mathsf{P}$. Note that although $\mathsf{P}_{\langle m(n)\rangle}$ is only defined on $\mathscr{F}_{n,m(n)}$ (cf. Proposition 30), we can extend it to $\mathscr{B}$ such that it agrees with the measure

P on every subset of $\mathbb{B}_n$ for each $n \geq 1$. Then for each $n \in \mathbb{N}$, we see that by setting $\mathbb{B}_n = \mathbb{D}^{\mathsf{C}}_{n,m(n)}$, both conditions i) and ii) in Lemma 39 are satisfied. From this and [31, Corollary 1′], we have the desired result.

## ACKNOWLEDGMENT

A. R. Iyengar would like to thank Aman Bhatia, Suhas Diggavi, Patrick Fitzsimmons, Henry Pfister, Bharath Sriperumbudur & Ruth Williams for stimulating discussions. The authors thank Eleni Drinea and Michael Mitzenmacher, respectively, for providing the numerical lower bounds for the BDC from [11] and the BRC from [5] (that appear as black circles in Figs. 1 and 2, respectively).

## REFERENCES

[1] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, "Modeling and information rate for synchronization error channels," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, 31 Jul.-5 Aug. 2011, pp. 518–522.
[2] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Problems Inform. Transmission*, vol. 3, no. 4, pp. 11–26, 1967.
[3] D. E. Knuth, "Big omicron and big omega and big theta," *SIGACT News*, vol. 8, pp. 18–24, April 1976. [Online]. Available: http://doi.acm.org/10.1145/1008328.1008329
[4] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. 2nd ed. Cambridge, Massachusetts: The MIT Press, 2001.
[5] M. Mitzenmacher, "Capacity bounds for sticky channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 72–77, Jan. 2008.
[6] ——, "A survey of results for deletion channels and related synchronization channels," *Probability Surveys*, no. 6, pp. 1–33, 2009.
[7] R. G. Gallager, "Sequential decoding for binary channels with noise and synchronization errors," *Lincoln Lab. Group Report*, 1961.
[8] S. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1226–1237, Mar. 2006.
[9] M. Mitzenmacher and E. Drinea, "A simple lower bound for the capacity of the deletion channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4657–4660, Oct. 2006.
[10] E. Drinea and M. Mitzenmacher, "Improved lower bounds for the capacity of i.i.d. deletion and duplication channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2693 –2714, Aug. 2007.
[11] A. Kirsch and E. Drinea, "Directly lower bounding the information capacity for channels with i.i.d. deletions and duplications," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 86 –102, Jan. 2010.
[12] Y. Kanoria and A. Montanari, "On the deletion channel with small deletion probability," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 13-18, 2010, pp. 1002–1006.
[13] A. Kalai, M. Mitzenmacher, and M. Sudan, "Tight asymptotic bounds for the deletion channel with small deletion probability," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 13-18, 2010, pp. 997–1001.
[14] S. Diggavi, M. Mitzenmacher, and H. D. Pfister, "Capacity upper bounds for the deletion channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 24-29, 2007, pp. 1716–1720.
[15] D. Fertonani and T. Duman, "Novel bounds on the capacity of the binary deletion channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2753 –2765, Jun. 2010.
[16] M. Rahmati and T. M. Duman, "Analytical lower bounds on the capacity of insertion and deletion channels," *CoRR*, vol. abs/1101.1310, 2011.
[17] Y. Kanoria and A. Montanari, "Optimal coding for the deletion channel with small deletion probability," *CoRR*, vol. abs/1104.5546, 2011.
[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. New York: John Wiley and Sons, 2006.
[19] M. Dalai, "A new bound on the capacity of the binary deletion channel with high deletion probabilities," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, 31 Jul.-5 Aug. 2011, pp. 410–413.
[20] D. Williams, *Probability with Martingales*. Cambridge University Press, 1991.
[21] A. R. Barron, "Limits of information, Markov chains, and projection," in *Proc. IEEE Int. Symp. Inf. Theory*, Sorrento, Italy, Jun. 25-30, 2000, p. 25.
[22] A. Feinstein, "On the coding theorem and its converse for finite-memory channels," *Information and Control*, vol. 2, no. 1, pp. 25–44, 1959.
[23] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.
[24] G. F. Lawler, *Introduction to Stochastic Processes*. Boca Raton, Florida: 2nd Ed. : Chapman & Hall/CRC, 2006.
[25] D. Blackwell, L. Breiman, and A. J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *The Annals of Mathematical Statistics*, vol. 29, no. 4, pp. pp. 1209–1220, 1958. [Online]. Available: http://www.jstor.org/stable/2236957
[26] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.
[27] D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498 –3508, aug. 2006.
[28] H. Pfister, J. Soriaga, and P. Siegel, "On the achievable information rates of finite state isi channels," in *Proc. IEEE Globecom 2001*, vol. 5, San Antonio, TX, USA, 25-29 Nov. 2001, pp. 2992–2996 vol.5.
[29] A. Kavcic, "On the capacity of Markov sources over noisy channels," in *Proc. IEEE Globecom*, vol. 5, San Antonio, TX, USA, Nov. 25-29, 2001, pp. 2997–3001.
[30] P. Vontobel, A. Kavcic, D. Arnold, and H.-A. Loeliger, "A generalization of the Blahut-Arimoto algorithm to finite-state channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1887 –1918, May 2008.
[31] R. L. Dobrushin, "Passage to the limit under the information and entropy signs," *Theory of Probability and its Applications*, vol. 5, no. 1, pp. 25–32, 1960.
[32] N. D. Vvedenskaya and R. L. Dobrushin, "The computation on a computer of the channel capacity of a line with symbol drop-out," *Problemy Peredachi Informatsii*, vol. 4, no. 3, pp. 92–95, 1968.
[33] J. Chen and P. Siegel, "Markov processes asymptotically achieve the capacity of finite-state intersymbol interference channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1295 –1303, march 2008.
[34] B. Kurkoski, P. Siegel, and J. Wolf, "Joint message-passing decoding of LDPC codes and partial-response channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1410 –1422, Jun. 2002.
[35] C. A. Desoer, "Communication through channels in cascade," Ph.D. dissertation, MIT, Cambridge, Massachusetts, 1953.
[36] R. Silverman, "On binary channels and their cascades," *IRE Trans. Inf. Theory*, vol. 1, no. 3, pp. 19–27, December 1955.
[37] M. Simon, "On the capacity of a cascade of identical discrete memoryless nonsingular channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. 16, no. 1, pp. 100–102, Jan 1970.
[38] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, "LDPC codes for the cascaded BSC-BAWGN channel," in *Proc. 47th Annual Allerton Conf. on Communication, Control and Computing*, Sep. 30 - Oct. 2, 2009, pp. 620–627.
[39] D. Fertonani and T. Duman, "Novel bounds on the capacity of binary channels with deletions and substitutions," in *Proc. IEEE Int. Symp. Inf. Theory*, 28 Jun.-3 Jul. 2009, pp. 2552 –2556.
[40] Z. Liu and M. Mitzenmacher, "Codes for deletion and insertion channels with segmented errors," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 224–232, Jan. 2010.
[41] F. Wang, D. Aktas, and T. M. Duman, "On capacity and coding for segmented deletion channels," in *Proc. 49th Annual Allerton Conf. on Communication, Control and Computing*, 28-30 Sep., 2011, pp. 1408–1413.
[42] R. Venkataramanan, S. Tatikonda, and K. Ramchandran, "Achievable rates for channels with deletions and insertions," *CoRR*, vol. abs/1102.5112, 2011.
[43] J. Hu, T. Duman, M. Erden, and A. Kavcic, "Achievable information rates for channels with insertions, deletions, and intersymbol interference with i.i.d. inputs," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1102 –1111, Apr. 2010.
[44] D. Fertonani, T. Duman, and M. Erden, "Bounds on the capacity of channels with insertions, deletions and substitutions," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 2–6, Jan. 2011.
[45] S. Shamai and E. Zehavi, "Bounds on the capacity of the bit-shift magnetic recording channel," *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 863 –872, May 1991.
[46] C. P. M. J. Baggen, "An information theoretic approach to timing jitter," Ph.D. dissertation, University of California San Diego, La Jolla, CA, 1993.

[47] D. Arnold, A. Kavcic, R. Kotter, H.-A. Loeliger, and P. Vontobel, "The binary jitter channel: a new model for magnetic recording," in *Proc. IEEE Int. Symp. Inf. Theory*, Sorrento, Italy, 25-30 Jun. 2000, p. 433.

[48] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, "Write channel model for bit-patterned media recording," *IEEE Trans. Magn.*, vol. 47, no. 1, pp. 35 –45, Jan. 2011.

[49] A. Mazumdar, A. Barg, and N. Kashyap, "Coding for high-density recording on a 1-d granular magnetic medium," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7403–7417, Nov. 2011.

[50] R. B. Ash, *Information Theory*. New York : Wiley, 1965.

[51] T. W. Benjamin, "Coding for a noisy channel with permutation errors," Ph.D. dissertation, Cornell University, Ithaca, NY, 1975.

[52] R. Ahlswede and A. Kaspi, "Optimal coding strategies for certain permuting channels," *IEEE Trans. Inf. Theory*, vol. 33, no. 3, pp. 310–314, May 1987.

[53] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150 –3165, Jul. 2008.

[54] L. Cui and A. Eckford, "The delay selector channel: Definition and capacity bounds," in *12th Canadian Workshop on Information Theory (CWIT)*, May 2011, pp. 15–18.

[55] V. Anantharam and S. Verdu, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.

[56] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. North Holland, 1977.

[57] S. I. Resnick, *A Probability Path*. Birkhauser Boston, 2005.